Measuring Behavior 2016

10th International Conference on Methods and Techniques in Behavioral Research
25-27 May 2016, Dublin, Ireland

Proceedings

Spink, Andrew
Riedel, Gernot
Zhou, Liting
Teekens, Lisanne
Albatal, Rami
Gurrin, Cathal (Eds.)

Proceedings of Measuring Behavior 2016

10th International Conference on Methods and Techniques in Behavioral Research

Volume Editors

Andrew Spink
Noldus Information Technology
E-mail: A.Spink@noldus.nl

Gernot Riedel
University of Aberdeen
E-mail: g.riedel@abdn.ac.uk

Liting Zhou
Dublin City University
E-mail: becky.zhou@dcu.ie

Lisanne E.A. Teekens
University of Twente
E-mail: lisanneteekens@gmail.com

Rami Albatal
Heystaks
E-mail: rami.albatal@dcu.ie

Cathal Gurrin
Dublin City University
E-mail: cathal.gurrin@dcu.ie

ISBN :978-1-873769-59-1

Preface

These proceedings contain the papers presented at Measuring Behavior 2016, the 10th International Conference on Methods and Techniques in Behavioral Research. The conference was organised by Dublin City University in cooperation with the University of Aberdeen and Noldus. The conference was held during May 25-27, 2016 in Dublin, Ireland.

We greeted the attendees at Measuring Behaviour 2016 with the following address: *Táimid an-bhrodúil fáilte a chur romhaibh chuig Baile Átha Cliath agus chuig an deichiú Comhdháil ar Iompar Tomhais. Tá súil againn go mbeidh am iontach agaibh anseo in Éirinn agus go mbeidh bhur gcuairt taitneamhnach agus sásúil. Táimid an-bhrodúil go háirithe fáilte a chur roimh na daoine ón oiread sin tíortha difriúla agus na daoine a tháinig as i bhfad i gcéin. Tá an oiread sin páipéar curtha isteach chuigh an chomhdháil seo go bhfuil caighdeán na bpáipéar an-ard ar fad agus táimid ag súil go mór le hócaid iontach"*. We are delighted to welcome you to Dublin for the 10th Conference on Behavioral Measurement. We hope that the you have a wonderful stay in Ireland and that your visit is both enjoyable and rewarding. We are very proud to welcome visitors from both Ireland and abroad and we are delighted to be able to include in the proceedings such high quality papers.

Building on the format that has emerged from previous meetings, we have a fascinating program about a wide variety of methodological aspects of the behavioral sciences. In addition to purely scientific presentations scheduled into nine oral sessions and ten symposia (covering a topical spread from rodent to human behavior). We have ten speakers in the demonstration showcase events, in which academics and companies demonstrate their latest prototypes. The scientific program also contains a workshop, two tutorials and a lunch-time session on scientific publishing.

We hope this program caters for many of your interests and we look forward to seeing and hearing your contributions and trust it will become a productive, exciting and memorable conference.

May 2016

Cathal Gurrin, Andrew Spink, Gernot Riedel
Chairs
MB2016

The Measuring Behavior Conference Series

Measuring Behavior is a unique conference about methods and techniques in behavioral research. While most conferences focus on a specific domain, Measuring Behavior creates bridges between disciplines by bringing together participants who may otherwise be unlikely to meet each other. At a Measuring Behavior conference, you find yourself among ethologists, behavioral ecologists, neuroscientists, experimental psychologists, human factors researchers, movement scientists, robotics engineers, software designers, electronic engineers, human computer interaction specialists to name but a few. Experience tells us that the focus on methodological and technical themes can lead to a very productive cross-fertilization between research fields. Crossing the boundaries between disciplines and species (from astronauts to zebras) can be extremely inspiring.

Measuring Behavior started in 1996 as a workshop in the framework of a European research project Automatic Recording and Analysis of Behavior, aimed at sharing the results of our project with colleagues from abroad. Organized by Noldus Information Technology and hosted by Utrecht University, Measuring Behavior 96 attracted over 150 participants. From that modest beginning, the conference has grown to a significant international event with several hundred delegates from thirty plus countries. This year is no exception with participants from more than thirty countries registered to attend at the time of going to print.

Over the years, the conference has been hosted by a variety of universities in many locations:

– 1996 (Utrecht) - Berry Spruijt
– 1998 (Groningen) - Jaap Koolhaas
– 2000 (Nijmegen) - Alexander Cools
– 2002 (Amsterdam) - Gerrit van der Veer
– 2005 (Wageningen) - Louise Vet
– 2008 (Maastricht) - Harry Steinbusch
– 2010 (Eindhoven) - Boris de Ruyter
– 2012 (Utrecht) - Remco Veltkamp, Gernot Riedel
– 2014 (Wageningen) - Gernot Riedel, Egon L. van den Broek, and Maurizio Mauri
– 2016 (Dublin) - Cathal Gurrin, Rami Albatal, and Gernot Riedel

In all previous years, the conference had been organised by Noldus Information Technology, who served as the conference organizer and main sponsor, with a number of additional sponsors every year. Measuring Behavior 2016 is the first time that the conference has taken place outside of the Netherlands and is primarily organised by Dublin City University and the Insight Centre for Data Analytics, together with Noldus IT. The organisers have made a big effort to put in place a compelling academic experience as well as an engaging social experience. The welcome reception at the Guinness Storehouse (Irelands number 1 tourist attraction and recently voted the best in Europe) is indicative of the focus on providing a memorable participant experience.

We have also grown in terms of the scientific quality of the conference, with selection of papers now being determined by a process of independent peer-review by an international team of hundreds of reviewers. The scientific program committee is very grateful for all that work that many of you reading this have contributed towards. In the scientific program, which is well balanced between human and animal research, you can find a variety of formats for presentation, interaction and exchange of information. In the past years we have seen that the symposia have become more prominent, and also the demonstration showcase has become more popular. Measuring Behavior is a scientific conference, so special attention is paid to publication of the work presented at the meeting. An important feature of the conference proceedings is that they are all available as open access from http://www.measuringbehavior.org. Now you find yourself at the 10th Measuring Behavior conference, the first conference to take place outside of the Netherlands. The organizers have done their best to prepare an optimal mix of scientific, technical, and social ingredients. We hope that you will find Measuring Behavior 2016 a rewarding and stimulating experience and wish you a pleasant stay in Dublin.

Organization

MB2016 is organized by the School of Computing, Dublin City University, The Insight Centre for Data Analytics, the University of Aberdeen and Noldus.

Organising Committee

Conference Chairs	Cathal Gurrin (Dublin City University, IE) Rami Albatal (HeyStaks, IE) Gernot Riedel (Univ, UK)
Program Chairs	Andrew Spink (Noldus IT, NL)
Organizing Chairs	Liting Zhou (Dublin City University, IE) Natasja Bogers (Noldus IT, NL) Lisanne E.A. Teekens (Noldus IT, NL, & University of Twente, NL)

Referee Panel

Rami Albatal (HeyStaks Technologies Ltd.)
Rasha Alblowi (Dublin City University)
Joseph Anthony (Dublin City University)
David Azcona (Dublin City University)
Melinda Babits (University of Debrecen)
Yoav Benjamini (Tel Aviv University)
Edda Bild (INCAS3 / University of Amsterdam)
Ryszard Bobrowicz (University of Copenhagen)
Natasja Bogers (Noldus)
Maarten Douwe Bredero (Architect Maarten Douwe Bredero)
Andre Brown (MRC Clinical Sciences Centre)
Nicolas Busquet (University of Colorado-Anschutz Medical Campus)
Ricardo Carbajo (Dublin City University / Insight)
Maurizio Casarrubea (University of Palermo)
Argyris Chatzitofis (SDS Games)
Vernica Mara Corrales-Carvajal (Champalimaud Centre for the Unknown)
Owen Corrigan (Dublin City University / Insight)
Anna Cranston (University of Aberdeen)
Barry Crouch (Aniversity of Aberdeen)
Wim Crusio (CNRS and University of Bordeaux)
Fabrice de Chaumont (Institut Pasteur)
Boris de Ruyter (Philips Research)
Ren de Wijk (WUR/FBR)
Serena Deiana (Boehringer Ingelheim)
Artur Direito (University of Auckland)
Ioannis Doumanis (Middlesex University)
Aaron Duane (Dublin City University / Insight)
Uli Eisel (University of Groningen)
Fiona French (London Metropolitan University)
Marcio Funes (University of So Paulo)
Ilan Golani (Tel Aviv University)
Huw Golledge (Universities Federation for Animal Welfare)
Alex Gomez-Marin (Instituto de Neurociencias de Alicante)
Shane Gore (Dublin City University / Insight)
Leonardo Gualano (Dublin City University / Insight)
Rashmi Gupta (Dublin City University / Insight)
Cathal Gurrin (Dublin City University / Insight)
Ehsan Habibi (Isfahan University of Medical Sciences)
Balzs Pter Hmornik (Budapest University of Technology and Economics)
Penny Hawkins (RSPCA)
Jer Hayes (IBM, Ireland)
Graham Healy (Dublin City University / Insight)
Tobias Heffelaar (Noldus Information Technology)
Marc Herrera (University of Barcelona)
Zaher Hinbarji (Dublin City University / Insight)
Ilyena Hirskyj-Douglas (University of Central Lancaster)
Julius Hodosy (Comenius University in Bratislava)
Frank Hopfgartner (University of Glasgow)
Feiyan Hu (Dublin City University / Insight)
Wolfgang Hrst (Utrecht University)
Iveel jargalsaikhan (Dublin City University / Insight)
Hvard Johansen (UIT ,The Arctic University of Norway)
Hideo Joho (University of Tsukuba)
Ida Kathrine Jrgensen (KRAK)
Rianne Kaptein (TNO)
Jochen Klein (Goethe University)
Agata Kokocinska (Institute of Genetics and Animal Breeding of the Polish Academy of Sciences)
Mt Kles (Budapest University of Technology and Economics)

Elisavet Ioanna Kyriakou (Radboud University / Noldus IT)
David Labbe (Nestl Research Center)
Zhang Li (Northeastern University)
Malte Thorben Lorbach (Noldus IT)
Andries Maat (Max-Planck-Institute for Ornithology)
Michel Mahieu (Janssen R&D BE)
Aleksas Mamkaitis (University of Limerick)
Clara Mancini (Open University)
Giuseppe Manfr (Radboud University Medical Centre)
Remy Manuel (Radboud University)
Francisco-Javier Martin-Arenas (UNIVERSIDAD DE SEVILLA)
Elke Mattheiss (AIT Austrian Institute of Technology)
Maurizio Mauri (IULM University of Milan)
Valeria Melis (University of Aberdeen)
Bent Egberg Mikkelsen (Aalborg University)
Alexander Mironov (St. Petersburg State University)
Eva Mohedano (Dublin City University / Insight)
David Samuel Monaghan (Dublin City University / Insight)
Pierre-Henri Moreau (University of Aberdeen/ TauRx)
Dongyun Nie (Dublin City University / Insight)
Kees Nieuwenhuis (Thales Nederland)
Anton Nijholt (University of Twente)
Lucas Noldus (Noldus IT)
Matthijs Noordzij (University of Twente)
Michael O'Mahony (University College Dublin)
Kwabena Ofei (Aalborg University)
Naohisa Ohta (Imagineering Institute)
Martin Ouwerkerk (Philips Group Innovation)
Cristina Palmero (Noldus IT / KU Leuven)
Fabio Patern (CNR-ISTI)
Christine Petr (Universit de Bretagne Sud)
Patricia Pons (Polytechnic University of Valencia)
Marie Postma-Nilsenova (Tilburg University)
Joo Prudente (University of Madeira)
Christopher Pryce (University of Zurich)
Vicen Quera (University of Barcelona)
Matthias Rauterberg (Eindhoven University of Technology)
Janet Reed (Texas AM University)
Carlos Ribeiro (Champalimaud Centre for the Unknown)
Gernot Riedel (University of Aberdeen)
T. Bas Rodenburg (Wageningen University)
Vincent Roy (Laboratoire PSY-NCA - EA4700)
Bahjat Safadi (LMSI)
Monica Sarret (University of Barcelona)
Philip Scanlon (Dublin City University / Insight)
Katrin Schenk (Randolph College)
Oliver Schreer (Fraunhofer Heinrich Hertz Institute)
Giulia Songa (Iulm university)
Andrew Spink (Noldus IT)
Matthew Stephen Sullivan (MMU)
Lisanne Teekens (Noldus IT / University of Twente)
Hans Theuws (Noldus IT)
Peter Tiernan (Dublin City University)
Michael Tsoory (Weizmann Institute of Science)
Emanuela Tullo (University Of Milan)
Wendie Uitterhoeve (MARIN)
Hansfried Van Craenendonck (Janssen)
Egon L. van den Broek (Utrecht University)
Ysbrand van der Werf (VU University Medical Center)
Eddy A. van der Zee (University of Groningen)
Lenny van Erp (HAS University of applied Sciences)
Erica van Herpen (Wageningen University)
Remco Veltkamp (Utrecht University)
Sarah Ellen Webber (University of Melbourne)
Haloin Wei (Dublin City University)
Andreas Weiglein (Medical University of Graz)
Craig Weiss (Northwestern University Medical School)
Michelle Westerlaken (Malmo University)
Yair Wexler (Tel Aviv University)
Hanna Wirman (Hong Kong Polytechnic University)
Fan Xu (The HongKong Polytechnic University)
Qishan Yang (Dublin City University)
Dian Zhang (Dublin City University / Insight)
Liting Zhou (Dublin City University / Insight)

Sponsoring Institutions

We acknowledge the significant input of our sponsors.

Table of Contents

Keynote Talks

Oral Papers: Fish

Oral Papers: Rodent Behaviour

Oral Papers: Precision Livestock Measurements

Oral Papers: Human Food & Diet 1

Oral Papers: Human Food & Diet 2

Oral Papers: Human Behaviour & Performance

Oral Papers: Human Measurement(psychology)

Oral Papers: Human Factors

Oral Papers: Analysis & General Methods

Posters

Symposia: Measuring Vocal Communication in Rodents

Symposia: The 9-5 Roden-Time for Change?

Symposia: Monitoring Neuronal Molecules During Behaviour Analytics?

Symposia: Measuring Choice Behaviour

Symposia: Combining Behaviour with Brain Imaging

Symposia: Analysis of Automatically Measured Behaviours of Humans and Animals

Symposia: Measuring Human Communication

Symposia: Human Behaviour Sensing and Technology Empowerment

Keynotes

From Small Sensors to Big Data: How the Sensor Web is Changing our World

Barry Smyth

University College Dublin
barry.smyth@insight-centre.org

our increasingly digitized world almost everything we do creates a data record that is stored somewhere, whether we are purchasing a book, calling a friend, ordering a meal, or renting a movie. And in the world of the sensor web this no longer limited to our online activities: exercising in the park, shopping for groceries, falling asleep, or even having a shower, are just some of our everyday activities that are likely to generate data.

The world of the sensor web is all about understanding how we can (and whether we should) use this information to enable better decisions. Better decisions for where we might live or where to send our kids to school. Better decisions about the food we eat and the exercise we should take. Better decisions by our governments and policy makers when it comes to managing education, energy, infrastructure, and healthcare. And better decisions for business and enterprise when it comes to understanding customer needs and demands.

In this talk we will discuss the origins of the sensor web and the attendant big data revolution. To understand where we will get to in the future we need to understand where we have come from in the past. But it is a unique feature of the present technological revolution that the pace of change is no longer stable and linear, rather it is accelerating and exponential. We will consider the implications of this when it comes to predicting how these technologies may develop over the coming years and decades. And we will explore how these developments point to entirely new ways of thinking about some of modern society's toughest challenges, and how the resulting technologies will impact on our everyday lives in the future.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Lifelogging — Personal Data for Human Measurement

Cathal Gurrin

Dublin City University, Ireland

Lifelogging is a new topic in data analytics and a form of pervasive computing which represents a phenomenon whereby people can digitally record their own daily live experience, in varying amounts of detail, for a variety of purposes. In a sense it represents the ultimate "black box" of a human's life activities and it offers huge potential as an assistive technology or as a means of recording or measuring aspects of an individual's lifestyle. We are currently observing a convergence of technologies which are creating the conditions for the emergence of lifelogging as a mainstream activity; computer storage has become cheap, there are readily available technologies for sensing the person, as well as location and environment and there is growing interest in the phenomenon of sensing and recording oneself, the quantified-self movement.

In this talk we will introduce and motivate the concept of maintaining lifelogs, which are set to revolutionise personal lives, healthcare, learning and productivity. Taking a human measurement viewpoint, we will motivate and describe the technical challenges that remain to be addressed and introduce the research activities aiming to address these challenges. We will introduce the technologies under development at various locations that attempt to efficiently gather flexible and extensible lifelogs and illustrate some early examples of their application in terms of human activity logging and measurement. We will end by describing the real-world experiences of lifeloggers who have been actively logging life experience for almost a decade.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Lacking quality in research: Is Behavioral Science Affected More Than Other Areas of Biomedical Science?

Anton Bespalov

Partnership for Assessment and Accreditation of Scientific Practice (PAASP), Heidelberg, Germany

When a clinical Phase II trial fails to meet its primary predicted endpoints, the preclinical data upon which the prediction was made is one area in the development chain that is often called into question. The reasons for translational failures are undoubtedly multifactorial but recent discussions increasingly focus on robustness of the preclinical data. Indeed, extensive analyses in several fields such as stroke, multiple sclerosis and amyotrophic lateral sclerosis confirm that preclinical efficacy data are not always as robust as claimed in the literature. These problems are not limited to any specific therapeutic area, academic or industrial research and are due to several generic factors influencing research quality (i.e. related to definition of pre-specified endpoints, principles of study design and analysis, biased reporting, and lack of proper training). Yet, partly driven by the fact that many current examples stem from the Neuroscience drug discovery, it is sometimes assumed that this area is affected more than the others. Further, there are concerns expressed about behavioral studies at risk of being poorly designed, underpowered and mis-reported. While there is no solid justification for such claims, it is important to review and analyze sources of bias and limited robustness that may be unique for behavioral pharmacology: multitude of environmental conditions that are difficult to control and that are often not reported (from housing conditions to subclinical infection); ethical concerns about in vivo research and the 3R pressure (as one of the major causes of studies being often under-powered); complexity of study design and analysis creating too much room for post hoc data „massaging“ and selective reporting; logistics of the study design and conduct (impact on blinding and randomization procedures); blood-brain barrier as a frequently neglected research tool quality factor; peculiar dose-effect functions („inverted U shape“) that are seldom viewed as a sign of poor study robustness. It is important to recognize all of these factors and take them into account when planning and conducting a study using behavioral methods in preclinical Neuroscience. It is argued that these basic improvements will greatly improve our ability to bring new innovative CNS medicines to patients.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Oral Papers

Description of a Multivariate Behavioral Test Arena for Zebrafish – the Zebrafish Multivariate Concentric Square Field Test

E. Roman[1], N. Tjernström[1] and S. Winberg[2]

[1]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden, Erika.Roman@farmbio.uu.se;
[2]Department of Neuroscience, Uppsala University, Uppsala, Sweden, Svante.Winberg@neuro.uu.se

Introduction

There is a strong need to develop new behavioral tests in order to gain progress in generating valid animal models for the understanding of pathophysiology in complex human disorders and in development of new treatment strategies [8].

The zebrafish, *Danio rerio*, possess many advantages as a model organism including low maintenance costs, short generation time and high homology to humans [4]. Many conventional behavioral tests for rodents [2] and zebrafish [5] have limited reproducibility and are intended for interpretation of specific predetermined mental states (e.g. "anxiety-like" behavior), which results in limited opportunities for more extended analyses of the various processes that presumably interact to establish complex behavioral patterns.

The multivariate concentric square field™ (MCSF) test for rats and mice, developed at Uppsala University, is unique in its design on the basis that the various areas should provide different elements of risk versus shelter and explorative incentives [6]. The guiding principle for the test is that it is unprejudiced, i.e. not designed for interpretation of specific predetermined mental states but instead enables a behavioral profiling of the animal. Thus, the MCSF performance is based on evolutionary conserved innate behaviors in a novel environment where areas of potential risk and shelter are incorporated [6, 7].

The aim of this study is to establish a MCSF test arena suitable for zebrafish. The novel tank diving (NTD) test was used as a reference test.

Material and methods

Male and female zebrafish from three different strains were used (n = 20 per sex and strain); AB, which is common lab strain, as well as wild-caught and pond-reared fish both originating from the Calcutta area, India. The experimental protocol and use of animals in this study was approved by the Uppsala Animal Ethical Committee, and was consistent with the Swedish Legislation on Animal Experimentation (Animal Welfare Act SFS1998:56) and the European Union Directive on the Protection of Animals Used for Scientific Purposes (2010/63/EU).

The fish were separated by sex 4 days before the experiments started. In this experiment, the fish were first tested in the NTD test followed by the MCSF test. They spent the night between the two tests in solitary tanks.

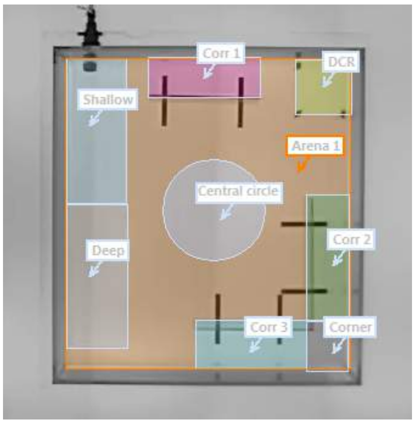


Figure 1. The MCSF arena with the different zones marked out.

The MCSF consists of a tank made of Plexiglas (50 × 50 × 30 cm; 13 cm water depth). The walls of the arena are sandpapered to prevent different cues in the room to affect the behavior. The different parts in the MCSF arena were a Plexiglas ramp with weights, a roof made of IR-transparent plastic, and three walls with weights. These parts were placed in the arena to form the various zones: dark corner room (DCR), central circle, corridor 1, corridor 2, corridor 3, corner and the ramp which was divided into deep and shallow (Figure 1). An IR-light table was placed underneath the arena and an IR-sensitive camera recorded the fish from above. The fish were placed in the arena with a net and allowed to freely explore for 30 minutes.

The NTD test consists of a rectangular Plexiglas tank (5 × 15 × 20 cm) divided into three zones; bottom, middle and top. The bottom and sides of the tank were covered with white plastic. An IR-sensitive camera was placed in front of the tank to capture the location and activity of the fish and an IR-light table was placed behind the tank. The fish was poured in to the tank from a transport beaker and a lid was placed on top. The testing time was 6 minutes.

The fish were tracked using Ethovision® XT 11.0 (Noldus Information Technology, Wageningen, The Netherlands). The number of visits, latency (s) to first visit, total time (s) sent, duration per visit (D/F, s) and distance travelled (cm) in each zone was registered, as well as mean velocity (cm/s) and distance travelled (cm) in the total arena during the trial.

Non-parametric statistics and multivariate data analysis (SIMCA-P+ 12.0, Umetrics AB, Umeå, Sweden) was used to examine MCSF and NTD performance. A principal component analysis was used to obtain an overview of the data, e.g. groups of observations, trends and outliers, and also to uncover the relationships between observations and variables.

Results

The chosen testing time (30 min) was enough for all fish to explore the MCSF arena, and individual differences in explorative strategies were evident. Moreover, using the zebrafish MCSF test we were able to detect clear differences in behavioral profiles between all three strains tested. As expected AB fish were more active, showed higher risk taking and were more explorative than fish of the two wild-type strains. We also detected significant differences in behavioral profile between males and females; males being more risk taking, interpreted as bolder, than females. This sex difference was most obvious in the AB strain.

The fish tested in the MCSF test were also monitored in the NTD, a more established behavioral assay for screening zebrafish for boldness. In general, the results on risk taking show good correspondence between the tests. Fish being more risk taking in the MCSF test (e.g. high activity in the central circle and on the shallow part of the ramp) were faster to explore the upper zone in the NTD test. However, it was also obvious that the divergence in behavioral profiles of the zebrafish strains tested was more clearly detected by the MCSF than by the NTD test.

Discussion

The results of the present study show that the zebrafish MCSF test is useful for screening for behavioral profiles. The MCSF test has the advantage that individual fish could be screened for multiple behavioral traits (e.g. general activity, risk taking, shelter seeking and exploratory activity) using a single test. In the current study we could show clear behavioral differences between the AB and the two wild-type strains. The AB strain has been kept in the lab for a large number of generations and an increase in boldness, i.e. more explorative and risk taking, has been reported in response to domestication in other teleosts [3]. We also detected clear behavioral differences between male and female zebrafish of the AB strain; males being bolder than females. This is in accordance with the results by Dahlbom et al. [1]. The zebrafish MCSF needs further validation but it clearly has a great potential in screening behavioral profiles in zebrafish.

Acknowledgements

Funding from the Facias Foundation (E.R) and the Swedish Research Council Formas through the Animal Health and Welfare ERA-net WinFish (S.W.) is gratefully acknowledged. The experiments were conducted with support from Uppsala University Behavioral Facility (UUBF).

References

1. Dahlbom SJ, Lagman D, Lundstedt-Enkel K, Sundström LF, Winberg S. Boldness predicts social status in zebrafish (Danio rerio). *PLoS One*, 2011;6: e23565.
2. Hånell A, Marklund N. Structured evaluation of rodent behavioral tests used in drug discovery research. *Front Behav Neurosci*, 2014;8: 252.
3. Huntingford F, Adams C. Behavioural syndromes in farmed fish: Implications for production and welfare. *Behaviour*, 2005;142: 1207-1221.
4. Kalueff AV, Stewart AM, Gerlai R. Zebrafish as an emerging model for studying complex brain disorders. *Trends Pharmacol Sci*, 2014;35: 63-75.
5. Maximino C, Herculano AM. A review of monoaminergic neuropsychopharmacology in zebrafish. *Zebrafish*, 2010;7: 359-78.
6. Meyerson BJ, Augustsson H, Berg M, Roman E. The concentric square field: a multivariate test arena for analysis of explorative strategies. *Behav Brain Res*, 2006;168: 100-113.
7. Meyerson BJ, Jurek B, Roman E. A rank-order procedure applied to an ethoexperimental behavior model - the multivariate concentric square field™ (MCSF) test. *J Behav Brain Sci*, 2013;3: 350-361.
8. Stewart AM, Kalueff AV. Developing better and more valid animal models of brain disorders. *Behav Brain Res*, 2015;276: 28-31.

Using a Plus Maze to Detect Effects of Geomagnetic Field Modifications on Fish Behavior

V.V. Krylov*, V.A. Nepomnyashchikh, N.A. Pankova, V.V. Pavlova and E.A. Osipova

I.D. Papanin Institute for biology of inland waters Russian Academy of Sciences, Borok, Russian Federation. * kryloff@ibiw.yaroslavl.ru

Abstract. Influence of geomagnetic field (GMF) on fish behavior is a well-established phenomenon. However, a variety of experimental devices have been used to study this phenomenon in fish making it difficult to compare results obtained by different researchers. On the other hand, plus-shaped maze is widely used as a standard device to study impacts of various factors on fish behavior. We have combined original data with the previously published ones to describe locomotor activity and orientation of roach larvae, juveniles and underyearlings, as well as adult zebrafish males, in the maze. Obtained results reveal that fish's behavior in the maze is sensitive to GMF and its disturbances. In addition, stereotypic movement patterns were discovered when fish visited arms of the maze. These patterns could obscure fish's orientation in GMF. We propose the plus-shaped maze as a suitable device to study the impact of GMF and its modifications on fish behavior, as well an interaction between responses to GMF and the aforementioned patterns. In addition, our results show that magnetic fields are to be taken into account when interpreting fish's directional preferences and locomotor activity in mazes and other experimental devices.

Introduction

The geomagnetic field (GMF) exerts various effects on fish behavior. First, fishes from different families are able to use the GMF for orientation [11]. For example, zebrafish, *Danio rerio*, prefer certain directions in GMF in a uniformly illuminated arena [21]. We also observed the magnetic orientation in adult zebrafish males using a plus-shaped maze with its arms oriented in N, W, E and S directions. Zebrafish's directional preferences were bimodal in the natural GMF: they visited E and W arms more frequently. This bimodal preference remained stable for same individuals across experiments divided by several days. When the horizontal GMF component was turned by 90 degrees clockwise, zebrafish's preference shifted accordingly by 90 degrees to N and S arms [16].

Second, locomotor activity in fish changes in response to GMF changes [11]. In zebrafish, the 90 degrees turn of GMF horizontal component resulted in a significant increase of fish's locomotor activity in comparison with the natural GMF. This increase became even more pronounced when the horizontal component was repeatedly turned by 90 degrees and back with 1 minute interval between turns [16].

Third, it was shown that an exposure of embryos to disturbed GMF could lead to changes in adult fish behavior. In a recent experiment by Romanovskij with co-authors zebrafish embryos were exposed to a simulated geomagnetic storm for 24 h. Adult zebrafish that have developed from these embryos started to explore a novel environment more quickly than those reared in the normal GMF [19].

Thus, various effects of GMF fluctuations on fish behavior are established phenomena. In this paper, we investigate the following problems regarding the study of these effects:

1. A variety of experimental devices had been used to study the effects of GMF on fish. This makes it difficult to compare results, obtained by different researchers. On the other hand, standard tests had been developed to study fish behavior, e.g., when comparing effects of various drugs. One of the most widespread tests consists in the monitoring of fish's behavior in a plus-shaped maze [see e.g.: 5, 7, 8]. We have already detected GMF effects on zebrafish's orientation and locomotor activity in such maze [16]. The results indicate that plus-shaped mazes may be one of standard tools to detect fish's

responses to changes in GMF. In order to verify the applicability of this approach for other fish species, we studied the orientation of roach, *Rutilus rutilus*, larvae, juveniles and underyearlings in GMF using a plus-shaped maze.

2. As an exposure of embryos to disturbed GMF could lead to subsequent changes in fish behavior, we, therefore, investigated, if these changes would be seen in a plus-shaped maze. To this end, roach embryos were exposed to distorted diurnal geomagnetic variation. Then larvae, juveniles and underyearlings raised from these embryos were tested in the maze.

3. When an animal is placed into an unfamiliar environment, the process of exploration follows a certain strategy [9 and refs. therein]. A lateralization of turning behaviour is also known for some fishes: when facing a choice between turning right- or leftward, they prefer one of these directions [1]. Potentially, the exploratory strategy and lateralization of turns in a plus-shaped maze could interfere with magnetic orientation and obscure it. We verified this hypothesis using data, obtained earlier in experiment on zebrafish [16] and new data on roach reported in this paper.

General methods

Ethical statement

The study was conducted in accordance to national and international guidelines (directive 2007/526/EC of the European Commission) for the protection of animal welfare. All experimental protocols and animals used in this research were approved by the Ethical Committee of I.D. Papanin Institute for Biology of Inland Waters Russian Academy of Sciences. All necessary permits were obtained for the field procedures with *R. rutilus*.

Apparatus

Fish were tested in plus-shaped mazes with four arms oriented to the North, East, South and West.The walls of maze were constructed of mat white LEGO blocks (LEGO Group, Denmark) and secured on the mat white bottom of rectangular plastic tank. The experiments were conducted in a dark room. An opaque cylindrical screen (diameter of 38 cm, height of 20 cm) made of white shaggy foam-rubber was placed around the maze. Fish cannot see experimenters or external landmarks due to the screen. The maze was lit by a ring of LEDs mounted 23 cm above the water surface. The light of LEDs was reflected and diffused by the foam-rubber, thus providing uniform illumination with no shadows within the maze.

Procedure

At the start of an observation, a fish was placed in the northern arm of maze which was isolated from the central square by a white plastic hurdle. One minute after that the hurdle was removed and fish's movements were recorded by a Panasonic-HC-X900M (Japan) video camera mounted 50 cm above the centre of maze.

Data analysis

Directional preferences and locomotor activity. We enumerated arms of maze from 1 to 4 clockwise when analyzing the videorecords, so a sequence of visits to arms was represented as a sequence of numbers 1, 2, 3 and 4 following each other in various order. A number of visits to different arms were counted. A visit was recorded if fish's head was inside an arm. Upon leaving an arm, fish almost never returned to it immediately (e. g., 22 or 33). If such a double visit happened, we counted it as a single one. In order to evaluate a preference of particular arms, we calculated a percentage of visits to each arm relative to the total number of all arms visits. The percentages had normal distribution (Shapiro–Wilk W-test, $p > 0.05$). A factorial ANOVA was performed to test for the influence of GMF and its modifications on fish preference to visit arms directed to N, E, S and W. The homoscedasticity assumption was satisfied (Levene's test).

The total number of visits to all arms was used to evaluate fish locomotor activity. The distribution of this parameter did not differ from the normal one, so the one-way ANOVA followed by Fisher LSD test as post-hoc test, was used to check for significance of differences in the locomotor activity between control and treated groups.

Patterns of visits to arms. Earlier it was shown, that zebrafish in a plus-shaped maze frequently visit arms following stereotyped patterns [18]. One pattern (shuttles) consists of repeating returns to already visited adjacent or opposed arm, for example, 12121212 or 242424. Another pattern (tours) is a repeating transition from one arm to another adjacent one, clockwise (1234) or in the opposite direction (3214). These patterns are interspersed by series of random visit sequences. The shuttles and tours are considered to be a part of fish's exploration strategy in a novel maze [18]. One may speculate that these movement patterns could interfere with the orientation in GMF. Therefore, we checked if fish in our experiments reveal such patterns. The following method was used. Let a sequence of visits to arms is 242423413... . The third, fourth and fifth visits are returns to already visited adjacent arms and designated as 1.The next visits are not returns and designated as 0. The resulting sequence for shuttles is **1110000 (** stands for first two visits which are not counted). Then we used the runs test and calculated *z* score [20]. If the score is significantly less than zero, then the 1s follow one another not randomly, but by series (shuttles). In a similar way, the above sequence of visits could also be checked for tours. In this case, seventh and eights visits are designated as 1, because they continue the clockwise motion which already have started at sixth visit. The resulting sequence for tours is **0000110.

Lateralization of turns. When fish leaves an arm and moves to an adjacent one, it should turn right or left. For each individual, the laterality index of the turns was calculated as $100\times(R-L)/(R+L)$, where R and L denote, respectively, the number of rightward and leftward turns of a fish. Thereafter, a mean laterality index for each trial was calculated over all fish individuals. The index varies from +1 to -1. Positive values of the index indicate a rightward lateralization of turns.

Roach

Methods

Fish and maintenance. Roe and sperm were collected from spawners caught in the Rybinsk reservoir (Yaroslavl region, Russia) during their natural spawning event in the first half of May, 2015. The method of dry fertilization was used [14]. Upon fertilization, eggs were randomly splitted into three groups (control and two experimental treatments, see below). Each group was placed in a separate jar filled with reservoir water (diameter of each jar 230 mm, height 75 mm, water level 60 mm), approximately 2500-3000 eggs per jar. The mean water temperature during embryo incubation was 17.5 ^{o}C. The temperature regime corresponded to the natural conditions found in the Rybinsk reservoir. The light/dark cycle was natural. The water in the jars was replaced twice a day.

On the 11-12th day after the fertilization samples of larvae from the control and both treatments were used in behavioral tests. Other larvae from the control and both treatments were placed in separate 72 L aquaria (approximately 50 individuals from each group per aquaria) and fed with an artificial diet for 1 month. Approximately 500 larvae from each group were placed into separate ponds and fed a natural diet for 4 months. Samples of 1 month old juvenile from aquaria and 4 month old underyearlings from ponds were also used in behavioral tests.

Exposure to GMF perturbations. Control embryos developed under a natural diurnal geomagnetic variation (the daily rhythm of changes in GMF magnitude). In the first experimental treatment (6h group) the variation was shifted relative to the natural one by 6 h (that is, midnight GMF values were shifted to morning in that case). In the second experimental treatment (12h group) the variation was shifted by 12 h (that is, midnight values were shifted to noon). The exposure to natural and shifted variations started immediately after the fertilization and lasted for 10 first days of development (until disappearance of yolk sacs and inflation of swim bladders).

The diurnal geomagnetic variation was shifted using two identical systems of Helmholtz coils described earlier [13]. The natural variation was recorded two days prior to the exposure. The ongoing natural diurnal geomagnetic variation was compensated within the Helmholtz coils, and shifted variations (6 h shift in one of the systems, and 12 h shift in another) based on the record were reproduced instead.

Behavioral tests. We roughly adjusted a size of maze to a size of fish (Table 1).The video records lasted for 15 min for larvae and juveniles. Underyearlings moved relatively slow, so they were recorded for 25 min. The mean water temperature during all records was 19°. No mortality was observed during the tests. Upon the end of the experiments all fish were transferred to aquaria for further growth.

Table 1. Length of roach and size of plus-shaped mazes (mm).

Age group	Length of roach (Mean ± S.E.)	Arm length	Arm width	Height	Water level
Larvae	7.99 ± 0.11	16	8	20	15
Juveniles	13.74 ± 0.22	24	16	20	15
Underyearlings	62.25 ± 0.87	110	95	80	70

Results

The mortality of *R. rutilus* embryos and hatching dynamics differed insignificantly among control, 6h and 12h groups.

Table 2. Locomotor activity (measured as a mean number of visits to arms) in roach.

Age group	Experimental group	Mean number of visits to all arms ± S.E.	Number of fish
	Control	114.77±15.35(a)	13
Larvae	6h	105.50±14.79(a)	14
	12h	164.44±13.84(b)	16
	Control	71.00±8.10(a)	19
Juveniles	6h	86.72±8.32(ab)	18
	12h	101.39±8.32(b)	18
Under	Control	79.30±10.86(a)	10
yearlings	6h	54.90±10.86(a)	10
	12h	84.71±12.98(a)	7

Note: Significant differences ($p < 0.05$) within each age group after LSD post-hoc multiple comparison tests are followed by different letters.

Table 2 shows the locomotor activity in roach, measured as the mean number of visits to all arms during 15 min in larvae and juveniles, and 25 min in underyearlings. The activity decreased as fish grew up. At the same time, the activity tended to be higher in 12h group in comparison to other groups of the same age. In larvae, the activity in 12h group was significantly higher than in control or 6h group. In juveniles, the significant differences were observed between 12h group and the control. On the other hand, no significant differences were revealed among groups of underyearlings.

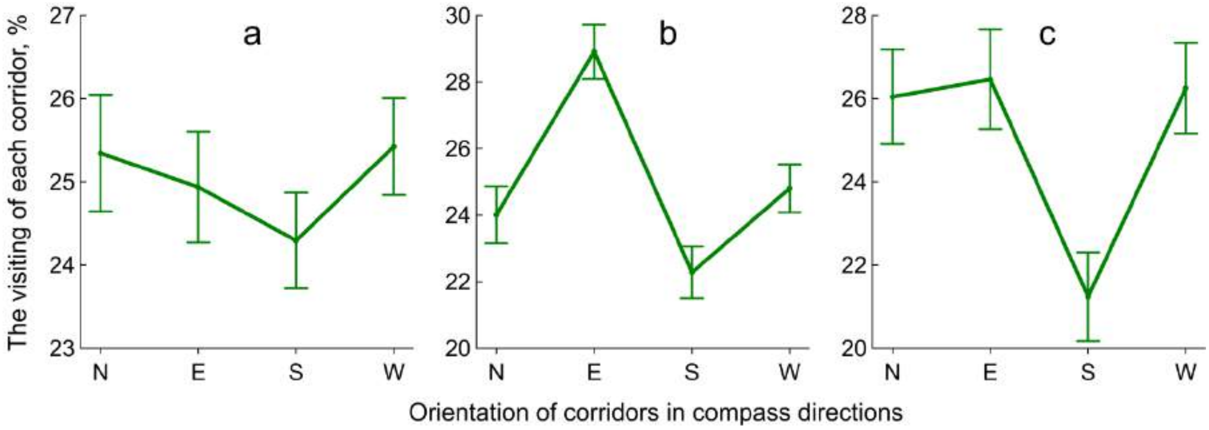


Figure 1. The preference of different arms in a plus maze derived from all experimental groups of larvae (a), juveniles (b) and underearlings (c). Whiskers show S.E.

ANOVA revealed no directional preference in any particular control, 6h, or 12h group. However, we observed a significant preference of E and W arms to N and S ones in all groups of juveniles taken together ($F_{3, 204}$=12.569, p<0.001) (Fig. 1). All groups of underyearlings taken together showed a significant preference of N, E and W arms to S one ($F_{3, 92}$=4.19, p<0.01). Larvae showed no significant preference at all ($F_{3, 160}$=0.648, p>0.05).

Table 3. Results of the runs test for roach.

Age group	Experimental group	Shuttles, mean *z* ± S.E.	Tours, mean *z* ± S.E.
	control	-1.18±0.45 (0.02)*	-2.20±0.46 (0.0004)*
Larvae	6h	-1.37±0.56 (0.03)*	-1.23±0.42 (0.01)*
	12h	-1.65±0.35 (0.0003)*	-2.50±0.49 (0.0001)*
	control	-1.32±0.24 (0.0001)*	-1.33±0.39 (0.004)*
Juveniles	6h	-1.18±0.34 (0.003)*	-1.23± 0.42 (0.009)*
	12h	-1.65±0.33 (0.0001)*	-1.97±0.41 (0.0002)*
Under	control	-1.06±0.31 (0.004)*	-1.91±0.33 (0.0001)*
yearlings	6h	+0.03±0.33(0.93)	-0.87±0.35 (0.04)*
	12h	-0.71±0.67 (0.33)	-0.69±0.49 (0.21)

Note: * *z* scores significantly differ from zero. Probabilities of error (p) are shown in parentheses.

Table 3 shows that mean *z* scores calculated for all groups of larvae and juveniles as well as for control underyearlings were significantly below zero (in fact, the scores were negative for a majority of individuals in these groups). The *z* scores calculated for tours were also significantly below zero in the 6h group of underyearlings. It does mean that roach's movements in these groups generally follow nonrandom stereotyped patterns: shuttles and tours. The insignificant results in other groups of underyearlings could be probably caused by a small number of fish.

No groups of roach prefer rightward or leftward direction of turns in the maze: mean laterality indices were close to zero and did not differ from it any significantly ($p>0.05$).

Discussion

The shifting of diurnal geomagnetic variation by 12 h during the early ontogenesis resulted in an increased locomotor activity of roach larvae and juveniles. The same tendency, albeit insignificant, was observed in underyearlings. One should take into account that the number of underyearlings used in the behavioral test was small, and the locomotor activity decreased as fish grew (see Table 2). These two facts possibly may explain the lack of a significant effect in underyearlings.

At the time being, one may only speculate on a possible mechanism of the effect. For example, it is believed, that diurnal geomagnetic variation is a secondary zeitgeber for biological circadian rhythms [17]. It was shown, that geomagnetic storms (in fact, disruptions of diurnal geomagnetic variation) influence a circadian rhythm of melatonin production in organisms [2, 3, 23]. Melatonin can provoke various effects on the development of central nervous system [4]. The circadian rhythm of melatonin production is observed in fish as early as on the second night after fertilization [10]. It is possible that distortions of diurnal geomagnetic variation in the present study lead to disruptions in circadian rhythm of melatonin production. This may cause changes in the nervous system and, therefore, behavior.

Juveniles and underyearlings preferred certain arms of the maze. As we conducted our tests in the uniformly illuminated apparatus with no visual cues, the observed result can be explained by orientation in GMF. Besides, it was shown earlier, that roach underyearlings are able to use GMF for orientation in a circular arena [12]. Larvae did not show any preference for particular arms. It is supposed that the majority of bony fishes use crystals of magnetite to detect an external magnetic field. The magnetite is accumulated in fish's tissues in the course of ontogenesis [15, 22]. It is possible that larvae have not accumulated enough magnetite for a magnetic orientation yet.

In addition to the orientation in GMF, roach tend to visit arms of plus-shaped maze following stereotypic patterns: shuttles between two arms and tours over different arms. Fish inevitably visited other arms besides E and W because of these patterns. Thus, stereotypic patterns interfered with the magnetic orientation. This interference explains, at least partially, why the observed preference for E and W arms was rather weak.

Zebrafish

Methods

Wild-type zebrafish (AB strain) were obtained from a commercial distributer, and then kept in laboratory for 2 months together in 70 L aquarium at 28^{o}C, 14:10 h light/dark cycle, and were fed daily with live Chironomid worms. Three days before the start of experiments the fish were placed into individual containers. During this period, the feeding with worms and the changing of water in containers were performed daily. 20 zebrafish males aged about 6 months were used for experiments. The standard length (i.e., distance from snout to base of caudal fin) was 26±3 mm.

Each arm of the maze was 60 mm long and 30 mm wide. The maze was 60 mm high, and filled with water up to 50 mm. The maze and experimental apparatus (see General Methods) were placed within the system of Helmholtz coils.

Fish movements were videotaped for 25 minutes. Then the fish was returned into its individual container. The half of males was tested first in the natural GMF. The next day they were tested with the horizontal GMF component turned 90 degrees clockwise from 5^{th} to 15^{th} min of the observation. The other half was tested in the reversed order [see ref. 16 for more details]. No mortality was observed during the tests. Upon the end of tests, zebrafish were returned to the aquarium.

Results

Zebrafish significantly preferred E and W arms of the maze in a natural GMF. The preference shifted to N and S arms during the period when the horizontal GMF component was turned 90 degrees clockwise. The directional preference was not 100%: non-preferred arms were repeatedly visited too both in natural and modified GMF. The locomotor activity increased during the turn of horizontal GMF component in comparison with natural GMF [16].

We used video records obtained by Osipova and co-authors [16] to calculate mean *z* scores for zebrafish movements in the natural GMF and turned horizontal GMF component. All the means, except the scores for the tours in natural GMF, were significantly below zero (Table 4). It does mean that zebrafish's movements do follows mainly nonrandom stereotyped shuttles and tours.

Table 4. Results of the runs test for zebrafish.

Experimental group	Shuttles, mean z ± S.E.	Tours, mean z ± S.E.
Natural GMF	-1.43±0.29 (0.0001)*	-0.69±0.33 (0.052)
The turn of horizontal GMF component	-1.29±0.34 (0.001)*	-0.99±0.33 (0.008)*

Note: * *z* scores significantly differ from zero. Probabilities of error are shown in parentheses.

Similar to roach, mean laterality indices were close to zero and did not differ from it any significantly ($p>0.05$). However, we found a significant correlation between individual laterality indices of fish in the natural GMF and indices of the same individuals in turned horizontal GMF component (Pearson's $r = 0.51$, $p<0.05$). It does mean that an individual fish have its own turning bias which is stable over repeated tests.

Discussion

Like roach juveniles, zebrafish showed a bimodal orientation in GMF: they preferred two opposite directions in GMF to other ones. This orientation changed when the horizontal GMF component was turned. These facts show that the observed directional preferences are accounted by magnetic orientation, and not by visual or other cues. A bimodal orientation in GMF was earlier observed in zebrafish [21]. It should also be mentioned that magnetite was recently revealed in zebrafish's tissues [6].

Similar to roach, zebrafish revealed stereotypic patterns of visit to arms: shuttles and tours. Therefore the patterns interfere with the magnetic orientation in zebrafish too. Besides these patterns, we observed the turning biases in zebrafish. It is possible that the biases could increase the fish's tendency to make tours, and thus contribute in an obscuring the magnetic orientation.

General discussion and conclusions

Both adult zebrafish males and roach, with the exception of pre-larvae, have demonstrated a preference for certain directions in the plus-shaped maze. These preferences can be explained by orientation in GMF.

Both zebrafish and roach juveniles visited other arms besides preferred ones. It seems logical to explain this fact by a stochastic nature of animal behavior as opposed to deterministic behavior of a machine. However, GMF is not the only factor which determines fish's behavior in the plus-shaped mazes. First, zebrafish and roach reveals stereotyped movements patterns, shuttles and tours which possibly are a part of their exploratory strategy in an unfamiliar environment. We suppose that the magnetic orientation and stereotypic patterns compete for the control of fish's movement direction. For example, tours interfere with the magnetic orientation: when fish visits every arm moving clockwise or counter clockwise, it obviously ignores the GMF direction. This competition may account for (at least, partially) the fact that fish visit non-preferred arms. Second, the orientation of arms to N, E, S and W does not necessarily coincide with preferences of a particular fish individual: it may prefer, for example, NE direction.

All this could be viewed as a drawback of the plus-shaped maze as a device for a study of precise magnetic orientation. However, an exploratory behavior in other devices could include the interference of stereotyped movement patterns too. Unlike a plus-shaped maze, such patterns are hard to detect in fish moving freely and following a wavy path in open arenas. An advantage of the plus-shaped maze is that fish should choose among few well defined directions. It provides an easier way to reveal stereotypic pattern of movements in such a device. Thus, the maze is useful for researchers who wish to investigate competition between the magnetic orientation and patterns of exploratory movements in fish. Besides, plus-shaped mazes have been shown as an efficient tool to study the impact of various factors on fish behavior (see Introduction). Therefore, we believe that it is reasonable to propose the use of these mazes as a candidate standard test for the study of interaction among responses to various factors on one hand, and stereotypic patterns of exploratory movements, on the other.

Finally, our results show that the effects of magnetic fields should be taken into account when interpreting fish's behavior. For example, influence of various drugs on zebrafish is measured as a change in the level of its locomotor activity and/or preference of dark or light arms in a plus-shaped maze [7]. Uncontrolled fluctuations of magnetic field in laboratory may distort results of these tests. Likewise, the possible competition of the GMF orientation and exploration strategy may obscure this strategy.

Acknowledgments

This work was partially funded by Russian Foundation for Basic Research (grants #15-04-06379-a and #16-34-00187-mol_a.). The research of behavioral lateralization was funded by Russian Science Foundation (grant #14-14-00284).

References

1. Bisazza, A., Pignatti, R., Vallortigara, G. (1997). Laterality in detour behaviour: interspecific variation in poeciliid fishes. *Animal Behaviour* **54**, 1273-1281.
2. Burch, J.B., Reif, J.S., Yost M.G. (1999). Geomagnetic disturbances are associated with reduced nocturnal excretion of a melatonin metabolite in humans. *Neuroscience Letters* **266,** 209-212.
3. Burch, J.B., Reif, J.S., Yost M.G. (2008). Geomagnetic activity and human melatonin metabolite excretion. *Neuroscience Letters* **438,** 76-79.
4. Chu, J., Tu, Y., Chen, J., Tan, D., Liu, X., Pi, R. (2016). Effects of melatonin and its analogues on neural stem cells. *Molecular and Cellular Endocrinology* **420**, 169-179.
5. Connors, K.A., Valenti, T.W., Lawless, K., Sackerman, J., Onaivi, E.S., Brooks, B.W., Gould, G.G. (2014).Similar anxiolytic effects of agonists targeting serotonin 5-HT 1A or cannabinoid CB receptors on zebrafish behavior in novel environments. *Aquatic Toxicology* **151**, 105-113.
6. Dixson, A. (2012). Zebrafish magnetite and long-lived rohon-beard neurons: Expanding our view of two zebrafish sensory systems in development and adulthood. *Ph.D.-Thesis*, California Institute of Technology.
7. Gould, G.G. (2011). Aquatic light/dark plus maze novel environment for assessing anxious versus exploratory behavior in zebrafish (*Danio rerio*) and other small teleost fish. In: Kalueff, A.V., Cachat, J.M. (Eds.), *Zebrafish Neurobehavioral Protocols*. Humana Press, New York, 99-108.

8. Grossman, L., Stewart, A., Gaikwad, S., Utterback, E., Wu, N., DiLeo, J., Frank, K., Hart, P., Howard, H., Kalueff, A.V. (2011). Effects of piracetam on behavior and memory in adult zebrafish. *Brain Research Bulletin* **85**, 58-63.
9. Kalueff, A.V., Zimbardo, P.G. (2007). Behavioral neuroscience, exploration, and K.C. Montgomery's legacy. *Brain Research Reviews* **53**, 328-331.
10. Kazimi, N., Cahill G. M. (1999). Development of a circadian melatonin rhythm in embryonic zebrafish. *Developmental Brain Research* **117**, 47-52.
11. Krylov, V.V., Izyumov, Yu.G.,Izvekov, E.I., Nepomnyashchikh, V.A. (2014). Magnetic fields and fish behavior. *Biology Bulletin Reviews* **4**, 222-231.
12. Krylov, V.V., Osipova, E.A., Pavlova, V.V., Batrakova, A.A. (2016). Effects of magnetic fields on spatial orientation in zebrafish *Danio rerio*(Cyprinidae) and roach *Rutilus rutilus* (Cyprinidae). *Journal of Ichthyology* (in press).
13. Krylov V.V., Zotov O.D., Klain B.I., Ushakova N.V., Kantserovaa N.P., Znobishcheva A.V., IzyumovYu.G., Kuz'mina V.V., Morozov A.A., Lysenko L.A., Nemova N.N., Osipova E.A. (2014). An experimental study of the biological effects of geomagnetic disturbances: The impact of a typical geomagnetic storm and its constituents on plants and animals. *Journal of Atmospheric and Solar-Terrestrial Physics* **110-111,** 28-36.
14. Lahnsteiner, F., Berger, B., Weismann, T. (2003). Effects of media, fertilization technique, extender, straw volume, and sperm to egg ratio on hatchability of cyprinid embryos, using cryopreserved semen. *Theriogenology* **60,** 829-841.
15. Mann, S., Sparks, N.H., Walker, M.M., Kirschvink, J.L. (1988). Ultrastructure morphology and organization of biogenic magnetite from sockeye salmon, *Onchorhynehusnerka*: implications for magnetoreception. *Journal of Experimental Biology* **140**, 35-49.
16. Osipova, E.A., Pavlova, V.V. Nepomnyashchikh, V.A. Krylov, V.V. (2016). Influence of magnetic field on zebrafish activity and orientation in a plus maze. *Behavioural Processes* **122**, 80-86.
17. Phillips, J.B., Deutschlander, M. (1997). Magnetoreception in terrestrial vertebrates: implications for possible mechanisms of EMF interaction with biological systems. In: *The Melatonin Hypothesis: Electric Power and the Risk of Breast Cancer*, 111-172.
18. Red'ko, V.G., Nepomnyashchikh, V.A., Osipova, E.A. (2015)/ Model of fish exploratory behavior in mazes. *Biologically Inspired Cognitive Architectures***13**, 9-16.
19. Romanovskij, A. V., Pesnya, D. S., Izvekov, E. I., Krylov, V. V., Nepomnyashchikh, V. A. (2014). The behavior of male Danio rerio Hamilton after exposure of fish embryos to a simulated geomagnetic storm. *Biophysics* **59,** 935-939.
20. Runyon, R. P. (1977). *Nonparametric statistics: a contemporary approach.* Addison Wesley Publishing Company, Reading, Mass.
21. Takebe, A., Furutani, T., Wada, T., Koinuma, M., Kubo, Y., Okano, K., Okano, T. (2012). Zebrafish respond to the geomagnetic field by bimodal and group dependent orientation. *Scientific Reports* **2**, 727.
22. Walker, M.M., Diebel, C.E., Haugh, C.V., Pankhurst, P.M., Montgomery, J.C., Green, C.R. (1997). Structure and function of the vertebrate magnetic sense. *Nature* **390**, 371-376.
23. Weydahl, A., Sothern, R.B., Cornelissen, G., Wetterberg, L. (2001). Geomagnetic activity influences the melatonin secretion at latitude 70 degrees N. *Biomedicine & Pharmacotherapy* **55 (Suppl.1),** S57-S62.

Development of a Fully Automated Operant System for Adult Zebrafish

Caroline H Brennan[1], Alistair J Brock[1], Matthew O Parker[2] and Bill Budenberg[3]

[1]Queen Mary University of London, United Kingdom
[2]Portsmouth University, United Kingdom
[3]Zantiks Ltd, Cambridge United Kingdom

Zebrafish are an established developmental genetic model species. Recently a number of high profile publications have suggested that zebrafish can also be used to explore the genetics and aetiology of behavior including complex behavioural phenotypes associated with psychiatric disease. However, progress has been limited. This is in part due to the lack of an automated system that can be used for analysis of complex behaviours and is suitable for large scale genetic or pharmaceutical screening. Here we describe our newly developed fully automated and scalable system for measuring a wide range of behaviors in adult zebrafish. The system is designed along the lines of a 'Skinner Box' for zebrafish. Stimuli are presented using a computer screen positioned beneath the holding tank. The tank has the dimensions 20x14x15 cm with the ability to include dividers to allow variations in internal space (see Figure. 1). The location of the fish is detected using an infra red camera positioned above the tank. The camera feeds in to a computer that drives the actuators to feed the fish or change the stimuli and also sends data to a linked network. The system is controlled by simple scripts that can be loaded on to the computer. Data is outputted as a csv file suitable for import into excel or any other data handling programme. We have used this system to establish methods for performing 2 - 5 choice discriminations based on colour or position of exemplars, to assess impulse control, learning and memory.

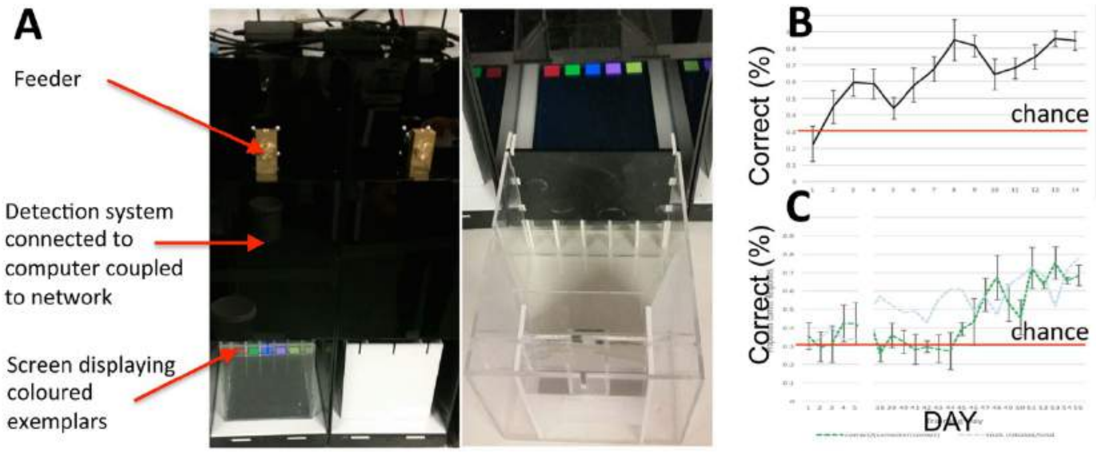


Figure 1. A) Automated operant system. Left hand panel; stimuli are presented using screens positioned on the base or sides (not shown) of the housing. A camera detects the fish and sends information to a computer that drives the actuators and sends data to the internet. Right hand panel tank with insert. B) 3 choice discrimination training data. C) 3 choice matching to sample data. Blue dashed line % choices made (max 30). Green line % correct ,

Following habituation to the tank and the feeding mechanism the training procedure has 3 stages: Step 1; train the fish to trigger food delivery by approaching a light (called the initiator light) adjacent to the food magazine. If the fish swims into a pre-defined domain in the vicinity of the light, the initiator light goes off and food is delivered. Both the light going off and the sound/vibration of the feeding motor signal imminent food delivery. Step 2; Fish initiate a trial by approaching the initiator light. Approaching the initiator light triggers the presentation of identical exemplars on the floor at the opposite end of the tank. The fish has to approach any of the exemplars to obtain food reward. Following a pre-determined time delay, when fish can eat the food, a second trial starts (the initiator light comes on). Step 3; in the third step the fish has to discriminate between the different exemplars. A correct choice is rewarded by food, an incorrect choice leads to the exemplars being turned off and a time out before the initiator light is turned on again signaling the start of another trial.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Typically fish are subject to 30 trials per day in a 30 min session. Performance is reported as percentage of initiated trials. Fish that do not consistently initiate at least 20 trials do not progress to stage 3. The use of this training scheme enables us to exclude fish that are unable to perform the task at an early stage ensuring that failure to discriminate between exemplars is an indication of discrimination ability rather than failure to learn the task per se. In addition to straight forward choice discriminations the system can be used to probe impulse control by varying the delay between the triggering of the initiator light and presentation of the choice lights. One can probe selective attention by adding distractors and memory by conducting matching to sample studies (for example by asking the fish to enter the choice chamber with the same colour as the initiator light). Using this system we have trained adult 4 month (at start) zebrafish to perform 3 choice discriminations to 80% accuracy within 3 weeks and 3 choice matching to sample to 60-70% accuracy within 8 weeks. This system allows the design of large scale genetic and drug screening projects to address complex behavioural phenotypes in adult zebrafish.

All animal experiments were approved by the QMUL Ethics Committee and performed under Home Office license in accordance with the Animals (Scientific Procedures) Act 1986, UK.

Correlation between Behavioral and EEG Components of Prenatal Hypoxed Rats

V.A. Vokina, L.M. Sosedova

Federal State Budgetary Scientific Institution East-Siberian Institution of Medical and Ecological Research, Angarsk, Russia. Vokina.vera@gmail.com

Introduction

At present, the interest in examining the relationship between the characteristics of prenatal development of the disease, not only during childhood but also in adults. Prenatal hypoxia is one of the major cause of mortality in newborns and cause of neurological disturbances in adulthood. The aim of this study was to assess behavioral, cognitive and electrophysiological effects of prenatal hemic hypoxia (PHH) during sensitive stages of brain maturation (E13 and E18) and to explore the relationships between this effects.

Materials and methods

The behaviour, spatial memory and EEG components was studied in naive rats and in rats with a model of sodium nitrite-induced hypoxia. Pregnant rats (n=20) were subcutaneously injected with 50 mg/kg sodium nitrite on 13 (E13) or 18 (E18) day of gestation. At the age of 3 months behaviors in offspring were assessed using open field and Morris water mase. The open field was a round arena of 1.2 m in diameter with radial and circular lines marked on the floor surrounded by 45 cm high walls. Four holes (2 cm in diameter) were cut into the floor of the arena. A video camera was mounted 1 m above the centre square. Bright overhead lighting was approximately 500 lux inside the arenas, while white noise was present at approximately 60 dB. Open field testing began approximately 4 days after toluene exposure. Animals were individually placed in the centre of arena and the following variables were recorded, by the use of a video camera, for 3 min: horizontal locomotor activity (numbers of crossed floor squares, movement time, duration of the immobility (resting time); vertical locomotor activity (number of rearing and climbing); number of entries in the central zone; number of hole visits; sniffing (number and duration); number of grooming, freezing and defecation [1]. The sequence and durations of behavioral patterns were recorded and analyzed using RealTimer (OpenScience, Moscow, Russia). The arena was carefully cleaned with alcohol and rinsed with water after each test, to removed environmental odours.The water maze was a circular 150-cm diameter swimming pool made of white PVC and filled 30-cm deep with 26 ±1 °C water. An 14-cm diameter white PVC platform was submerged 1.5 cm under the water surface level. The rats were brought to a waiting room at least 30 min before the experiments and were kept in holding cages shortly before testing. Within 2 days each animal 4 times placed in a water maze and registered latency to the platform (escape latency). Once on the platform, it was left there for 30 s. Between the trials, the animals were kept in holding cages for 60 s. After four trials each day, the rats were dried with paper towels and heated by two 75-W light bulbs.

After behaviour and spatial memory was measuring rats were anesthetized (ketamine 75 mg/kg and xylazine 10 mg/kg, i.p.) and mounted in a stereotaxic frame. Stainless steel electrodes (diameter 100 µm, length 2 mm) were implanted according to Paxinos and Watson's stereotactic atlas: 1.5-2 mm posterior to bregma и 2 mm lateral to midline (sensorimotor cortex); 1,5 mm posterior to lambda and 2 mm lateral the midline (visual cortex); 3 mm anterior to the bregma (reference). At least 1 week was allowed for the animals to recover from the surgery. The power spectra analysis of signals was obtained for delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta 1 (13–22 Hz) and beta 2 (22-32 Hz) frequencies. Registration of visual and auditory evoked potentials (VEP and AEP) was performed using a multifunctional complex for EEG and EP "Neuron-Spectrum-4" (Neurosoft, Russia). Latencies P1, N1, P2, N2 components and interwave amplitude N1P2 and P2N3 components was analyzed.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Statistical analysis was performed using Statistica software package (StatSoft version 6.1, № AXXR004E642326FA). The correlation between behaviour, spatial memory and EEG components was determined by Spearman's correlation test.

Results

The changes of behaviour and EEG components were peculiar to each period PHH. Prenatally (E13) nitrite exposed rats exhibited decreased total locomotion in the open field and increased latency to the platform in water maze unlike hypoxed E18 rats. Spectral analysis showed that prenatal (E13) nitrite treatment resulted in a reduction in background amplitude, increasing in EEG beta 1 and beta 2 power in visual and sensorimotor cortex. Prenatal (E13) hypoxed rats demonstrated significant increased latencies N1, P2, N2 and P3 components of VEPs and AEPs ($p<0.01$) without any changes interwave amplitudes of N1P2 and P2N2 component. Change in the index of basic EEG rhythms and VEP and AEP was not observed in rats prenatally (E18) hypoxed. Next, we verified whether the locomotion activity and spatial memory and EEG changes were correlated. In control rats and prenatally (E18) nitrite exposed rats was found statistically significant correlation between locomotion and the background amplitude of the visual cortex ($r= 0,77$; $p = 0.04$). The relationship between spatial memory and EEG in these groups was not detected. We found that number of correlations increased in the group of prenatally (E13) nitrite exposed rats. Thus, established a statistically significant positive relationship between locomotor activity and power δ-rhythm of the visual cortex ($r= 0,86$; $p = 0.01$). The negative correlations was found between locomotor activity and r θ-, α-, β1- β2- rhythms ($r> 0,71$; $p <0.05$); between freezing and θ-rhythm ($r= -0,63$; $p = 0.09$ in visual and $r= -0,73$; $p = 0.038$ in sensorimotor cortex, respectively).

A larger percentage of pairs with a correlation coefficient more than 0.7 (>90%) was detected in groups of prenatal hypoxed animals, leads to the conclusion that the stability of the formed functional state of the central nervous system. Required for normal functioning of CNS balance is achieved using the adaptive mechanisms and ability of the organism to maintain this balance depends on stage of brain development, on which there is negative impact.

Ethical statement

All aspect of the care and treatment of laboratory animals were treated according to the European Convention for the Protection of Pet Animals (Strasbourg, 1987). The animals were kept in single breeding cages under standard conditions (and provided with food and water ad libitum.

References

1. Sestakova, N., Puzserova, A. Kluknavsky, M., Bernatova I. (2013). Determination of motor activity and anxiety-related behaviour in rodents: methodological aspects and role of nitric oxide. *Interdisciplinary Toxicology*6(3),126–135.

Use Of Image Processing Techniques To Quantify Sensory and Motor Behaviors in Rodents By Measuring Whisker Movements

R.A.. Grant and B.M Hewitt

Conservation, Evolution & Behaviour Research Group, Manchester Metropolitan University, Manchester, UK. Robyn.grant@mmu.ac.uk

Introduction

Generating quantifiable metrics for the assessment of animal health and welfare is often intrusive (such as taking physiological samples), and places the animal under a great deal of stress [1]. Measurable behavioral models have been proposed as a way to mitigate a large part of the stress, and also allow for the animal to be studied while behaving freely. Several behavioral models have been developed in the past, including beam balancing (to test balance), rotorod (to test balance, coordination, stamina), novel object (to test cognition and memory) and morris water maze (to test memory and navigation). These tests require extensive animal training, and usually only measure the duration and frequency of certain behaviors. Most laboratories will expose their animals to a host of these behavioral tests in order to capture a range of behaviors; however, these approaches lack sufficient sensitivity to wholly characterize the progressions and recovery of animal health problems over different periods [2]. Indeed, there is still a priority in the Biological Sciences to develop and validate new measures, especially incorporating measurements of animal behavior, perception and cognition [3].

One particular quantitative behavior that has recently garnered a lot of attention is that of whisking [4,5]. Rodent whiskers, specifically, are regularly studied as a model system for investigating fundamental principles of sensory processing [6,7]. However, the vibrissal system is also specialized for high-speed motor control and during exploration, the whiskers can move forwards and backwards at rates of up to 25 Hz. With the recent development of high-speed filming techniques, it has become clear that whiskers can make very complex movements especially during object exploration. Indeed, rodents have the capacity to alter the timing, spacing and positioning of their whiskers to maximize sensory information [4,5] using a complex array of facial muscles. Although the novel muscle architecture in the whisker pad may play a role in resistance to aging and/or degenerative disorders [4], we aim to demonstrate here that whisker movements are affected in rodent models of motor disease, including motor neuron disease [4], aging, Parkinson's Disease and Huntington's Disease.

Methods

This study uses a range of mouse models with motor disorders, including Motor Neuron Disease and Aging (Sod1), Parkinson's Disease (SNCA-OVX) and Huntington's Disease (CAG250 R6/2). We compare the whisker behaviors of these animals to age-matched controls over the time course of the disease. All procedures were approved by the local Ethics Committee and UK Home Office, under the terms of the UK Animals (Scientific Procedures) Act, 1986. Mice were introduced to an open field Perspex arena (20 x 30 x 15 cm) (Fig. 1a), which was lit from below by a bright, infra-red light box (PHLOX LEDW-BL-400/200-SLLUB-Q-1R-24V). Mice were filmed from above using a digital high-speed video camera (Phantom Miro ex2) recording at 500 frames per second with a shutter-speed of 1 ms and a resolution of 640x480 pixels. Multiple 700 msec video clips were collected opportunistically (by manual trigger) when the animal moved in to the cameras field of view. Approximately 15 clips were collected from each animal. In each selected clip, the mouse snout and whiskers were tracked using the BIOTACT Whisker Tracking Tool (Perkon et al., 2011) (Figure 1d), which semi-automatically finds the orientation and position of the snout, and the angular position (relative to the midline of the head) of each identified whisker. From processing the angular position traces (Figure 1b,c) from the whiskers on each side of the head, we can generate average approximations of whisker frequency, amplitude, offset and speeds.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

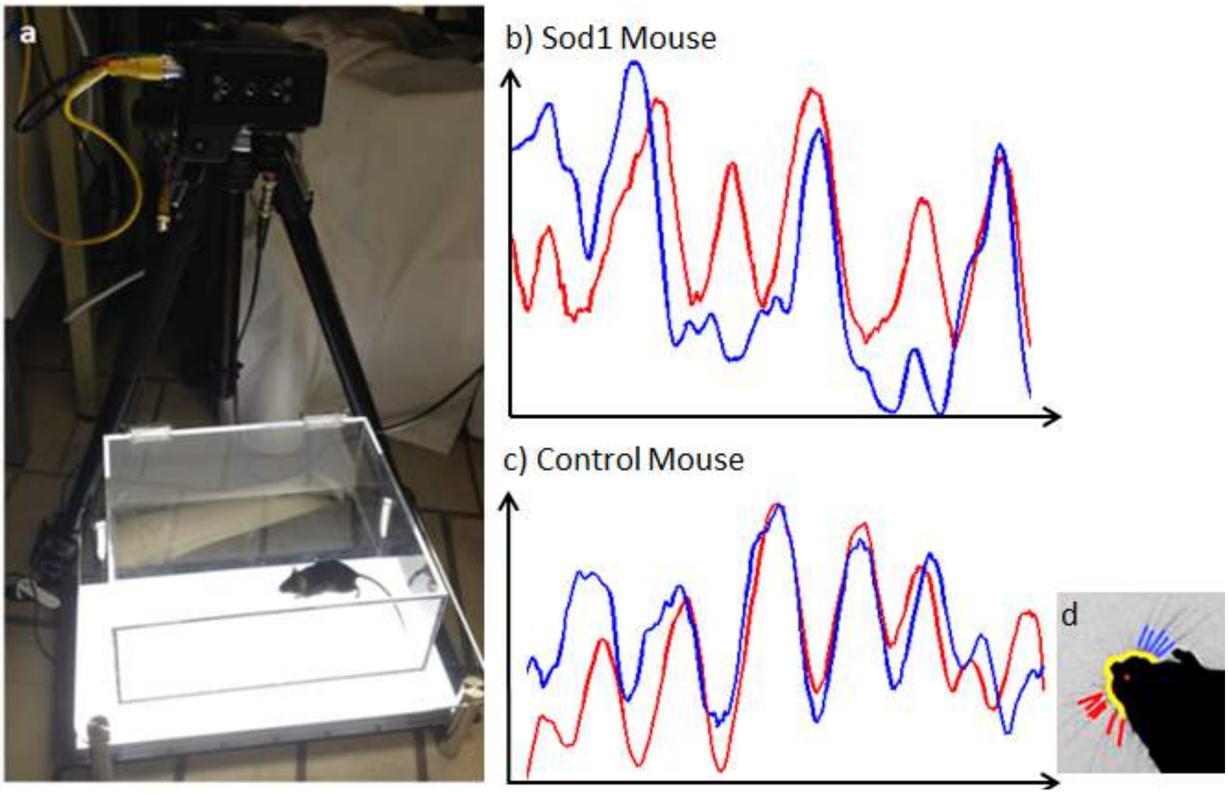


Figure 1. Photograph of behavioural set-up (a), with average whisker traces for the right (in blue) and left (in red) hand-side whiskers in a Sod1 (b) (model of Motor Neuron Disease), and control (c) mouse. Example whisker and head tracking can be seen in d.

Results

We observed that retraction speeds and amplitudes were both generally disrupted in mice with motor disorders. Specifically, in the Sod1 mice, we found that amplitude and retraction speeds were elevated in 120 day old animals, while whisker frequency and offset were both reduced at this time-point (Figure 1b,c). In the SNCA-OVX mice we found that amplitude and retraction velocity were both reduced at 18 months. While the significant changes in whisking behavior have only been observed towards the later stages of the disease in both the Sod1 and SNCA-OVX mice, in the CAG250 R6/2 we also saw robust changes earlier in the disease course. Specifically, we observed that CAG250 R6/2 mice had a period of hyper-motion at 10 weeks, including increased whisker protraction and retraction speeds, and elevated offsets and amplitudes. This occurred for less than two weeks, and by 18 weeks whisker movements were substantially reduced, with low retraction speeds and amplitudes.

Discussion

We suggest that our novel arena set-up, and image processing software that simultaneously measures locomotion and whisker movements, provides robust and quantitative behavioral measurements of object exploration and motor control. That whisking is easily quantified, and intrinsically linked to object exploration, cognition and perception in rodents [5,8], makes it a good candidate to address the strategic priority to develop new behavioral measures. The next stage in developing our set-up is to incorporate a pedobarograph and associated software, in order to measure gait parameters alongside whisker movements. This will allow for comprehensive quantification of motor behaviors in rodents, by integrating both locomotion and whisker analyses.

References

1. Tieu, K. (2011). A guide to neurotoxic animal models of Parkinson's disease. *Cold Spring Harbor perspectives in medicine*, *1*(1), a009316.
2. Casarrubea, M., Davies, C., Faulisi, F., Pierucci, M., Colangeli, R., Partdrige, L., ... & Benigno, A. (2015). Acute Nicotine Induces Anxiety and Disrupts Temporal Pattern Organization of Rat Exploratory Behavior in Hole-Board: A Potential Role for the Lateral Habenula. *Frontiers in Cellular Neuroscience*, *9*, 197.
3. BBSRC Strategic Priorities. Accessed 7 December 2011.
4. Grant, R. A., Sharp, P. S., Kennerley, A. J., Berwick, J., Grierson, A., Ramesh, T., & Prescott, T. J. (2014). Abnormalities in whisking behaviour are associated with lesions in brain stem nucle i in a mouse model of amyotrophic lateral sclerosis. *Behavioural brain research*, *259*, 274-283.
5. Arkley, K., Grant, R. A., Mitchinson, B., & Prescott, T. J. (2014). Strategy change in vibrissal active sensing during rat locomotion. *Current Biology*, *24*(13), 1507-1512.
6. Diamond, M. E., von Heimendahl, M., Knutsen, P. M., Kleinfeld, D., & Ahissar, E. (2008). 'Where' and 'what' in the whisker sensorimotor system. *Nature Reviews Neuroscience*, *9*(8), 601-612.
7. Kleinfeld, D., Ahissar, E., & Diamond, M. E. (2006). Active sensation: insights from the rodent vibrissa sensorimotor system. *Current opinion in neurobiology*, *16*(4), 435-444.
8. Mitchinson, B., & Prescott, T. J. (2013). Whisker movements reveal spatial attention: a unified computational model of active sensing control in the rat. *PLoS Comp Biol*, *9*(9), e1003236.

Train the Sedentary Brain: Burrowing as a Novel Strength Exercise Method in Mice

P. Roemers[1], P.N. Mazzola[1], H.F Godthelp[1], P.P. De Deyn[2], W.J. Bossers[3] and E.A. van der Zee[1]

1 Molecular Neurobiology, Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, The Netherlands
2 Department of Neurology and Alzheimer Research Center, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
3 Center for Human Movement Sciences, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Background

In both mice and men, physical activity increases brain power and brain health. Studies in humans and rodents do not only show that exercise improves cognitive functioning, but also show that exercise can counter cognitive decline in the context of aging and neurodegenerative diseases. [1,2] Research in this field has focused on the effects of endurance exercise, or aerobic exercise, for the most part. This is especially true in rodent studies, where use of voluntary running in a running wheel or forced running on a treadmill has become the gold standard. However, a recent study performed in dementia patients indicates that a combination of strength exercise and aerobic exercise is more effective in slowing dementia related cognitive decline than aerobic exercise alone. [3]
It is unknown what type of strength exercise and aerobic exercise is the best combination to boost brain health or which duration of either type of exercise is optimal. Moreover, despite a wealth of data on the mechanisms underlying the effects of endurance exercise on the brain [4], it is unknown which brain mechanisms underlie the beneficial effects of strength exercise on the brain. This can be studied in mice. However, few voluntary strength exercise methods are available for mice. Resistance wheel running is used as a form of voluntary resistance exercise, but this type of exercise may have an aerobic component too. [5,6] Other strength exercise methods for rodents include hanging exercise or loaded ladder climbing and weight pulling on a treadmill. However, these methods are all forced exercise methods. [7,8]

A method which induces voluntary strength exercise has benefits when compared to forced exercise methods in experimental set-ups which assess learning and memory. Forced methods depend on handling the animal and using positive or negative reinforcement (i.e. shocks) to induce exercise behaviour. As a consequence, these forced methods are relatively stressful. Stress is known to be detrimental to the brain, and notably the hippocampus. [9] Voluntary exercise methods avoid the possibly confounding effects of stress and are therefore more suited to use in combination with behavioural tests, including cognitive tests.

Objective

The aim of this study is to test burrowing as a voluntary strength exercise method for mice.

Design/methods

Pilot experiments – We adapted and modified the burrowing apparatus and protocol that was described by Deacon. [10-12] (See Figure 1A-C) To test different set-ups for a burrowing protocol that could be used as an exercise method, we performed two pilot experiments which each lasted four weeks. We tested wood-chips, sand, pebbles and a mix of sand and pebbles as burrowing substrates in a pseudo-randomized block design, using adult male C57Bl6 mice (n=8). It was reported that mice may lose motivation to burrow over repeated burrowing sessions [12], so we tested a protocol in which mice had an opportunity to burrow every other day, as well as one in which the mice had access to a burrowing tube for five nights a week. Using this latter protocol, we recorded and scored burrowing behaviour by placing infra-red lighting and a night-vision camera above the cage to determine the amount of time spent burrowing. Moreover, we weighed the amount of substrate remaining in the tube after 1h, 3h or ~12h. The five days per week protocol was performed using pebbles and a mix of pebbles and sand only.

Main experiment - To test burrowing as a voluntary strength exercise method, male C57Bl6 mice, 10 weeks old, were used (n=10-12 per group). Mice had access to a burrowing tube five nights a week, for a period of fourteen weeks. The tube was filled with a mix of pebbles and sand (>1000 g) or, for pseudo-burrowers, with wood chips (>100 g).

In parallel, mice that had access to an unloaded or loaded running wheel served to test the effects of these already established methods of endurance and resistance exercise. Custom build resistance running wheels were used. The resistance of these wheels can be modified by adjusting the load which hangs from a pulley connected to the axis of the wheel. (See Figure 1C-F)

Motor coordination was measured in a balance beam test. Muscle strength was measured using a grip strength test before the exercise interventions and after five and ten weeks of exercise. Muscle stamina (or fatigability) was measured in the inverted screen holding test at the same time points. A progressive endurance treadmill test was used to measure endurance after twelve weeks of exercise. In week fourteen, mice were sacrificed and multiple forelimb and hindlimb muscles were weighed and frozen. Epididymal fat pads were weighed. Blood serum, cerebrospinal fluid and brain tissue were stored for future analyses. (See Figure 2 for an overview of the timeline)

Repeated measures ANOVA with pre-defined contrasts were used to assess the effect of exercise type on balance beam performance, grip strength and stamina (in the inverted screen test). One-way ANOVA was used to test for differences in endurance and tissue weights. All procedures were in accordance with the regulation of the ethical committee for the use of experimental animals of the University of Groningen, The Netherlands (License number DEC 6974A)

Results

Pilot experiments - The pilots show that most of the burrowing takes place in the first hour, or rather first half hour, as indicated by the frequency and duration of burrowing bouts, and the amount of substrate removed by the mice in the first hour as compared to the amount removed in the remaining hours of the dark phase. In the first hour, mice burrowed up to 900 grams of pebbles and sand out of a total of ~1000 grams overnight.

Main experiment - Mice will burrow ~1 kg of mix overnight. Again, most of the burrowing took place in the first hour. Burrowing was performed readily over fourteen weeks. Motor coordination on the balance beam was improved in both runners and resistance runners. Forelimb grip strength declined with age, but this decline was attenuated in both burrowers and resistance runners. Although burrowers did show the highest latency to fall, no significant differences between groups were found in the inverted screen holding test. Resistance running and running increase endurance in the treadmill test, as well as the maximum running speed. Both forms of running induced hypertrophy of the soleus muscle, which largely consists of aerobic muscle fibres. The most important results are summarised in Table 1.

	Functional changes			**Tissue weight**		
Goups (n=12)	Grip strength (/body weight)	Stamina (inverted screen)	Endurance (treadmill)	Soleus muscle (/body weight)	Other muscles * (/body weight)	Fat pad (/body weight)
Runners	=	=	▲(~40%)	▲(~20%)	=	▼(~35%)
Resistance Runners	▲ (~18%)	=	▲(~30%)	▲(~31%)	=	▼(~35%)
Burrowers	▲ (~13%)	=	=	=	=	=

Table 1 - Effects of burrowing, resistance running and running, as compared to home cage controls (n=10) and, in the case of the burrowers, to pseudo burrowers (n=12) as well. ´´▲´´increased, ´´▼´´ decreased or ´´=´´ unchanged as compared to

relevant controls. *Other muscles which were dissected and weighed are the Tibialis anterior, Quadriceps femoris, Gastrocnemius and Triceps brachii muscle.

Conclusion/Discussion

Based on our behavioural tests, both burrowing and resistance running are suitable methods for strength exercise: Both methods increase forelimb strength as compared to aged-matched control groups. Resistance running increases both strength and endurance, whereas burrowing increases strength only. This suggests burrowing may be more specific as a strength exercise method, whereas resistance running has an aerobic component too. Follow-up studies in progress will further characterize these methods and their effects on muscles and brain. Future studies will investigate the applicability of burrowing as a strength exercise method in aged mice and mouse models for Alzheimer's disease.

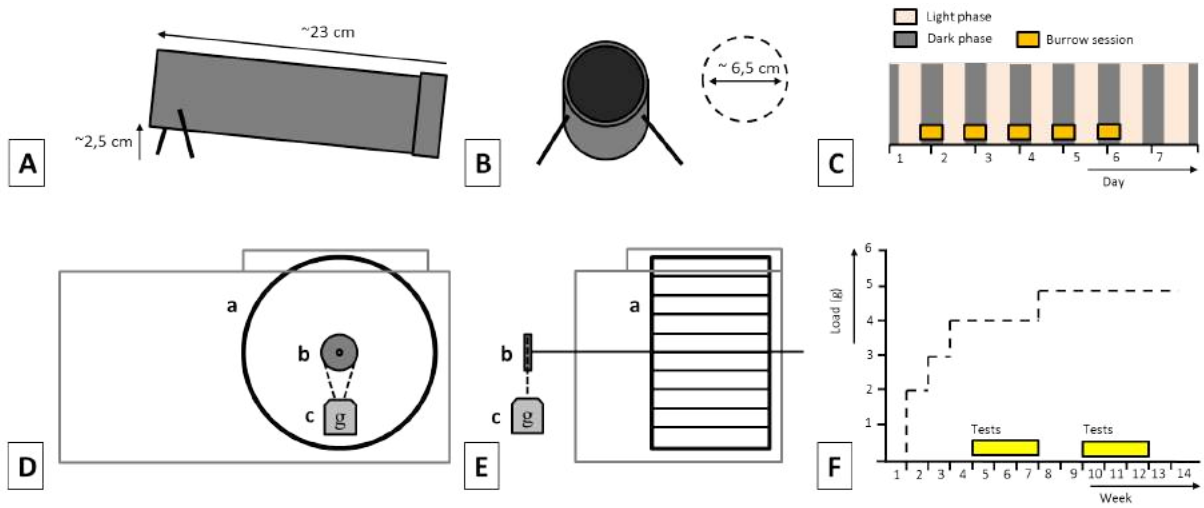


Figure 1 - Burrowing tube and resistance running wheel. The burrowing tube viewed from the side (A) and from the front (B). C) In the main experiment, the burrowing tube, filled with a mix of sand and pebbles, was placed in the cage of a mouse during the dark phase for five days each week. D-E) The resistance running wheel (a) had a pulley (b) connected to the elongated axis. Various weights (c) could be hung underneath the pulley to increase resistance on the wheel. F) The wheels were calibrated by hanging a small weight in the wheel at the height of the axis and adjusting the weight underneath the pulley until the wheel moved. The load was increased up to 5 grams by the end of the experiment.

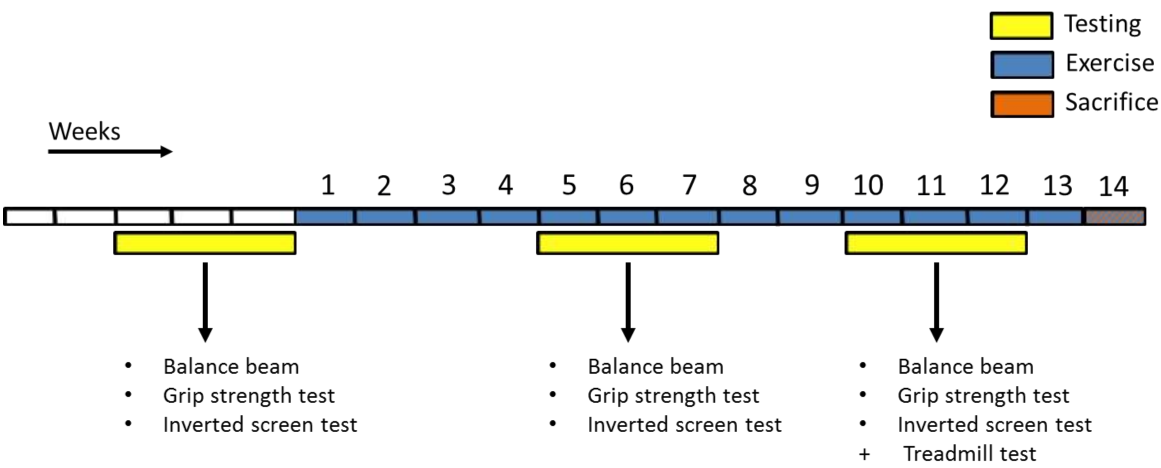


Figure 2 - Timeline of the main experiment in weeks. Behavioral tests were conducted to determine functional effects of running, resistance running and burrowing. After sacrifice, muscle tissue and fat pads were weighed. Muscles, brain tissue, blood serum and cerebrospinal fluid were collected and frozen.

References

1. Hillman, C. H., Erickson, K. I. & Kramer, A. F. (2008). Be smart, exercise your heart: exercise effects on brain and cognition. *Nat. Rev. Neurosci.* 9, 58–65.
2. Intlekofer, K. A. & Cotman, C. W. (2013). Exercise counteracts declining hippocampal function in aging and Alzheimer ' s disease. *Neurobiol. Dis.* 57, 47–55.
3. Bossers, W. J. R. *et al.* (2015). A 9-Week Aerobic and Strength Training Program Improves Cognitive and Motor Function in Patients with Dementia: A Randomized, Controlled Trial. *Am. J. Geriatr. Psychiatry*. doi:10.1016/j.jagp.2014.12.191
4. Voss, M. W., Vivar, C., Kramer, A. F. & van Praag, H. (2013). Bridging animal and human models of exercise-induced brain plasticity. *Trends Cogn. Sci.* 17, 525–44.
5. Konhilas, J. P. *et al.* (2005). Loaded wheel running and muscle adaptation in the mouse. *Am. J. Physiol. Heart Circ. Physiol.* 289, H455–65.
6. Soffe, Z., Radley-Crabb, H. G., McMahon, C., Grounds, M. D. & Shavlakadze, T. (2015). Effects of loaded voluntary wheel exercise on performance and muscle hypertrophy in young and old male C57Bl/6J mice. *Scand. J. Med. Sci. Sports* 1–17. doi:10.1111/sms.12416
7. Krüger, K. *et al.* (2013). Functional and muscular adaptations in an experimental model for isometric strength training in mice. *PLoS One* 8, e79069.
8. Seo, D., Lee, S., Kim, N. & Ko, K. Humanized animal exercise model for clinical implication. (2014). *Pflügers Arch.* 466, 1673–1687.
9. Kim, J. J., Song, E. Y. & Kosten, T.A. (2006). Stress effects in the hippocampus: synaptic plasticity and memory. *Stress* 9, 1–11.
10. Deacon, R. M. J. (2006). Burrowing in rodents: a sensitive method for detecting behavioral dysfunction. *Nat. Protoc.* 1, 118–121.
11. Deacon, R. M. J. (2009). Burrowing: A sensitive behavioural assay, tested in five species of laboratory rodents. *Behav. Brain Res.* 200, 128–133.
12. Deacon, R. (2012). Assessing Burrowing, Nest Construction, and Hoarding in Mice. *Journal of Visualized Experiments*. doi:10.3791/2607

Longitudinal analysis of voluntary and evoked behavioral paradigms in chronic neuropathic pain and the effect of social isolation

C. Pitzer[1], R. Kuner[2] and A. Tappe-Theodor[2]

[1]**Interdisciplinary neurobehavioral core (INBC), University of Heidelberg, Heidelberg, Germany**
[2]**Pharmacology Institute, University of Heidelberg, Heidelberg, Germany. anke.tappe-theodor@pharma.uni-heidelberg.de**

Background

For decades, rodent models have been used to investigate diverse clinically-relevant pain conditions. It is generally agreed that we need these models to investigate mechanisms and to develop novel treatments [1], although currently used rodent pain models are often criticized for not fully reflecting clinical pain characteristics [2-4]. Chronic pain is mostly of spontaneous nature and pain-attacks affect sociability and often the ability for voluntary behavioral tasks. These aspects are severely under-investigated in rodents. While patients can describe their pain verbally, most rodent studies rely on short-duration stimulus-evoked unilateral hindpaw measurements, even though chronic pain is usually a day and night experience.
Many preclinical studies are therefore limited and cannot represent the full pain picture. Moreover, various other aspects influence behavioral experiments. Among these are the presence and sex of a human observer [5] and the gender of the animals [6]. Other factors like over-handling or physical and social enrichment [7, 8] are widely discussed. Applying too many tests or restraining of the animals can lead to stress [9] and thereby stress-induced analgesia [10] or stress-induced hyperalgesia [11]. Furthermore, social isolation harbors stress conditions and can affect the pain behavior [12, 13]. Laboratory mice are often kept individually, because of specific intervention or problems with fighting when using male animals. These conditions are not considered sufficiently in most studies.

It is time to comprehensively and extensively characterize longstanding models and assess further changes in pain-related daily-life wellbeing [14]. Especially longitudinally measurements of voluntary behavior in unrestrained animals are missing and might provide important aspects for better bench-to-bedside translation. Moreover, we lack standardization in most behavioral paradigms to guarantee reproducibility and ensure adequate discussion between different studies. Important aspects are herewith the combination and duration of particular behavioral tasks and the effects of social isolation.

Aims

We aimed at thoroughly investigating stimulus-evoked and voluntary behavioral parameters in the Spared Nerve Injury (SNI) model for neuropathic pain [15]. Our emphasis was to assess voluntary behavioral parameters in an observer independent manner and over 24 hours, enabling the investigation of the circadian rhythm. Moreover, we analyzed the impact of social isolation in all behavioral paradigms.

Methods

We used a portfolio of classical response measurements for mechanical and temperature stimuli, detailed gait analysis, using two different measuring systems (Dynamic weight bearing system and Catwalk system), as well as observer-independent voluntary wheel running and home cage monitoring (Laboras system). These measures were performed on a weekly basis up 12 weeks following surgery. Voluntary wheel running and home cage monitoring was performed over 24 hours each time.

Results

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

SNI mice developed significant and stable mechanical allodynia and cold hyperalgesia throughout the observation period. We did not observe any significant difference between grouped and isolated SNI mice in the degree and extend of mechanical allodynia and cold hyperalgesia. SNI mice gained significantly less weight than sham mice and all isolated mice gained less weight than grouped mice.

We measured very robust changes in static gait parameters like paw print area and paw intensity contact in all SNI cohorts which were congruent with the duration of hyperalgesia and allodynia. Interestingly, a variety of dynamic gait parameters, like stance duration and duty cycle also temporally matched these changes, whereas some dynamic gait parameters were changed in a time-dependent manner. SNI mice of all cohorts showed an increase in the swing phase during the first two weeks following surgery and later a decrease in stride length.

We were very surprised that SNI animals did not show any significant reduced wheel running activity compared to their basal running activity. Nevertheless, we observed significant difference in wheel running activity between SNI and sham mice of the grouped cohort. Interestingly, voluntary wheel running behavior was not affected in SNI mice, but by social isolation:
Using the home cage monitoring system we measured significant differences at the day of operation between SNI and sham animals for most behavioral parameters in grouped and isolated animals. Besides a reduced climbing activity, SNI mice did not showe tremendous alterations in the home cage activity.

Conclusion

This is the first longitudinal study providing detailed insights into various voluntary behavioral parameters related to pain and highlights the importance of social environment on particular voluntary behavioral parameters in a mouse model of chronic neuropathy. From the above described data, it is clear that social isolation has an influence on the animal behavior with respect to weight gain and the motivation for voluntary tasks. The SNI model does not implicate any alteration in voluntary wheel running activity or home cage activity. On the contrary, we observed time-dependent changes in dynamic weight bearing parameters in the gait analysis and this highlights the importance of this behavioral paradigm in studying chronic pain elicited in the SNI model.

All procedures were in accordance with the ethical guidelines imposed by the local governing body (Regierungspräsidium Karlsruhe, Germany).

References

1. Mogil JS, Davis KD, Derbyshire SW: (2010) The necessity of animal models in pain research. *Pain* **151**:12-17.
2. Mogil JS, Graham AC, Ritchie J, Hughes SF, Austin JS, Schorscher-Petcu A, Langford DJ, Bennett GJ: (2010) Hypolocomotion, asymmetrically directed behaviors (licking, lifting, flinching, and shaking) and dynamic weight bearing (gait) changes are not measures of neuropathic pain in mice. *Mol Pain* **6**:34.
3. Urban R, Scherrer G, Goulding EH, Tecott LH, Basbaum AI: (2011) Behavioral indices of ongoing pain are largely unchanged in male mice with tissue or nerve injury-induced mechanical hypersensitivity. *Pain* **152**:990-1000.
4. Cobos EJ, Ghasemlou N, Araldi D, Segal D, Duong K, Woolf CJ: (2012) Inflammation-induced decrease in voluntary wheel running in mice: a nonreflexive test for evaluating inflammatory pain and analgesia. *Pain* **153**:876-884.
5. Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH, Wieskopf JS, Acland EL, Dokova A, Kadoura B, Leger P, Mapplebeck JC, McPhail M, Delaney A, Wigerblad G, Schumann AP, Quinn T, Frasnelli J, Svensson CI, Sternberg WF, Mogil JS: (2014) Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat Methods* **11**:629-632.
6. Sorge RE, Mapplebeck JC, Rosen S, Beggs S, Taves S, Alexander JK, Martin LJ, Austin JS, Sotocinal SG, Chen D, Yang M, Shi XQ, Huang H, Pillon NJ, Bilan PJ, Tu Y, Klip A, Ji RR, Zhang J, Salter MW, Mogil JS. (2015) Different immune cells mediate mechanical pain hypersensitivity in male and female mice. *Nat Neurosci* **18**:1081-1083.

7. Jirkof P, Cesarovic N, Rettich A, Nicholls F, Seifert B, Arras M. (2010) Burrowing behavior as an indicator of post-laparotomy pain in mice. *Front Behav Neurosci* **4**:165.
8. Pham TM, Hagman B, Codita A, Van Loo PL, Strommer L, Baumans V. (2010) Housing environment influences the need for pain relief during post-operative recovery in mice. *Physiology & behavior* **99**:663-668.
9. Mogil JS. (2009) Animal models of pain: progress and challenges. *Nat Rev Neurosci* **10**:283-294.
10. Terman GW, Shavit Y, Lewis JW, Cannon JT, Liebeskind JC. (1984) Intrinsic mechanisms of pain inhibition: activation by stress. *Science* **226**:1270-1277.
11. Imbe H, Iwai-Liao Y, Senba E. (2006) Stress-induced hyperalgesia: animal models and putative mechanisms. *Front Biosci* **11**:2179-2192.
12. Bravo L, Alba-Delgado C, Torres-Sanchez S, Mico JA, Neto FL, Berrocoso E. (2013) Social stress exacerbates the aversion to painful experiences in rats exposed to chronic pain: the role of the locus coeruleus. *Pain* **154**:2014-2023.
13. Vachon P, Millecamps M, Low L, Thompsosn SJ, Pailleux F, Beaudry F, Bushnell CM, Stone LS. (2013) Alleviation of chronic neuropathic pain by environmental enrichment in mice well after the establishment of chronic pain. *Behav Brain Funct* **9**:22.
14. Tappe-Theodor A, Kuner R. (2014) Studying ongoing and spontaneous pain in rodents-challenges and opportunities. *The European journal of neuroscience* **39**:1881-1890.
15. Decosterd I, Woolf CJ. (2000) Spared nerve injury: an animal model of persistent peripheral neuropathic pain. *Pain* **87**:149-158.

Cognitive “Omics”: Validation of Potential Drug Targets on the Basis of Cognitive Patterns

I. Gyertyán

MTA-SE NAP B Cognitive Translational Behavioural Pharmacology Group
Department of Pharmacology and Pharmacotherapy, Semmelweis University
Institute of Cognitive Neuroscience and Psychology, Research Center for Natural Sciences, MTA, Budapest, Hungary, gyertyan.istvan@med.semmelweis-univ.hu

Abstract

A realistic estimation for the number of potential cognitive enhancer mechanisms already identified in animal tests can be around a few hundreds. Yet, there are only 2 types of drugs in clinical application for dementia and memory impairment: the acetyl-choline-esterase (AchE) inhibitors and an NMDA antagonist (memantine). The increasingly tense unmet need has driven enormous R&D activity in the field, however, the clinical development of new drugs faced 100% attrition rate in the last decade, the main reason being lack of efficacy in Phase 2 and 3 trials [1,2]. The factors underlying this considerable translational gap are manifold, but low translational value of animal models is regularly brought into focus [3,4].

Molecular targets coming out from basic research need rigorous validation in relevant animal models to check their suitability for industrial drug development projects. If an animal model is intended to be predictive for the human situtation it should model as close as possible the human cognitive task and should conform to the human terminology. Impairing elementary memory functions (passive avoidance, novel object recognition) by a single dose of e.g. scopolamine, may well not be a valid model of the profound cognitive deficits characterizing the human disease. More complex paradigms with better face validity (e.g. attentional set-shifting) should be used and the cognitive deficit spectrum of the target disease, the various cognitive domains should be covered by an appropriate number of relevant assays. The recently elaborated MATRICS [5] or CNTRCS [6] systems are good examples for this approach.

Perhaps, the most critical step is modelling cognitive *deficits* (i.e. deteriorating cognitive functions). Defective cognitive performance can be brought about by several means: pharmacological agents, cerebral lesions/activations (including optogenetics), stressors, modulation of gene expression, old age, increasing task difficulty or selecting low performers of the population – all may yield low cognitive outcome amenable for improvement. Nevertheless, in lack of exact knowledge on or deliberately being unrelated to the pathomechanism of the given disease (symptomatic treatment) no distinct impairing intervention can be considered as the most „appropriate” or „predictive” (*see* e.g. the failure of the scopolamine-induced amnesia models in predicting clinical efficacy). This is a problem of construct validity. On the other hand, each type of impaired cognitive state may bear a certain extent of relevance to the human cognitive deficit. Consequently, the more types of impairing methods against which the studied mechanism is effective the higher the chance it will be effective against the cognitive defects (of unknown origin) of the target disease.

Thus, multiple types of cognitive impairment is suggested and instead of the actually fashioned „gold standard” cognitive assay, a ’cognitive domains x impairing methods’ matrix of models as test battery is proposed to be used for clinical prediction

Recently, we started to build up such a system, where several cognitive tasks (representing different cognitive domains) are taught to the same set of animals. For example, animals learn the five-choice serial reaction time task (for attention), attentional set-shifting task (rule learning), a modified Morris water maze task (episodic memory), rotarod performance (procedural memory). Hereby we create a population with widespread „knowledge” and the effect of a particular impairment method on the various cognitive functions can be simultaneously measured in this population. Presumably, the different cognitive domains will show differential

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

sensitivity for a given intervention. Further, by applying longer term impairments (i.e. subacute/subchronic drug treatment or stress) a „disease state" can be better mimicked; and the various impairing methods may bring about various „disease states" differing in the pattern of cognitive deficiencies. Potential cognitive enhancer mechanisms can then be tested in the distinct disease states. Here also, long term treatment is desirable to model the clinical situation. Presumably, the different cognitive deficits will show differential sensitivity for a given intervention. Using the same „patient population" provides a coherent background for efficacy comparison and cognitive enhancer pattern recognition and eliminates the „noise" on the results originating from methodical diversity found in the literature.

The resulting outcome of testing a particular cognitive enhancer mechanism in the above system is a cognitive enhancer pattern. This pattern may determine the clinical studies with a compound. Cognitive deficits of neurological and psychiatric disorders show diverse patterns [7] and this pattern-specificity may require molecules with different modes of action. The disease whose cognitive deficit pattern best matches the cognitive efficacy pattern of the selected molecule/mechanism should be chosen as the target clinical indication. Giving a simplistic example, if a compound shows outstanding efficacy in assays measuring attention then it should be tried in ADHD, whereas if it is more active in social cognition paradigms then autism could be the preferred choice. This pattern-matching approach, which we call cognitive „*omics*" because of the analogous logic, will likely increase the clinical success rate of the potential cognitive enhancers.

Nevertheless, as the pattern-matching approach implies elevated requirements for a certain mechanism or compound for being deemed „efficacious", the number of real hits will foreseeably be reduced. It may be considered good news on one hand, as the basic translational problem was the high number of false positive hits in the animal literature. On the other hand, it is also foreseeable that because of the scarcity of compounds capable to enter clinical trials and the longer preclinical investigation periods the industrial investors and top management may easily become disappointed and decide to refrain from CNS drug development – as it already happened in the near past. Experience has taught us that the described target validating experimentation does not fit the industrial R&D timeframe and scenery. Therefore, it should be done in the precompetitive area and out of the conventional industrial settings. Once the target is (deeemed to be) valid, drug discovery screening may return to its traditional way (back to the companies' R&D labs) and can be carried out in simple(r) assays with sufficient robustness and capacity. However, before entering into developmental phase with the optimized molecule, the selected clinical candidate should be double-checked in (some of) the target validating paradigms

Ethical statement: the animal experiments involved in the test system described by the paper were authorized by the regional animal health authority in Hungary (resolution number PEI/001/3572-4/2014) and conformed to the Hungarian welfare law and the EU 63/2010 Directive.

Acknowledgement

This study was funded by the Hungarian National Brain Research Program (NAP), contract # KTIA_NAP_13-2014-0015.

References

1. Kola, I., Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery,* **3**, 711-715.
2. Cummings, J.L., Morstorf, T., Zhong, K. (2012). Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimer's Research & Therapy,* **6**, 37-43.
3. Gartner, J.P. (2014). The significance of meaning: why do over 90% of behavioral neuroscience results fail to translate to humans, and what can we do to fix it? *ILAR Journal*, **55**, 438-456.
4. Steckler, T. (2015). Preclinical data reproducibility for R&D—the challenge for neuroscience. *Psychopharmacology* **232**, 317–320.
5. Young, J.W., Powell, S., Risbrough, V., Marston, H.M., Geyer, M.A. (2009). Using the MATRICS to guide development of a preclinical cognitive test battery for research in schizophrenia. *Pharmacology and Therapeutics*, **122**, 150–202.

6. Moore, H., Geyer, M.A., Carter, C. (2013). Harnessing cognitive neuroscience to develop new treatments for improving cognition in schizophrenia: CNTRICS selected cognitive paradigms for animal models. *Neuroscience and Biobehavioral Reviews*, **37**, 2087–2091.
7. Millan, M.J. Agid, Y., Brüne, M., Bullmore, E.T., Carter, C.S., Clayton, N.S., Connor, R., Davis, S., Deakin, B., DeRubeis, R.J., Dubois, B., Geyer, M.A., Goodwin, G.M., Gorwood, P., Jay, T.M., Joëls, M., Mansuy, I.M., Meyer-Lindenberg, A., Murphy, D., Rolls, E., Saletu, B., Spedding, M., Sweeney, J., Whittington, M., Young, L.J. (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature Reviews Drug Discovery,* **11***,* 141-168.

Virtual Environment Training in Mice as a Naturalistic Platform to Enable Operant Behaviour

M. Havenith[1,2], P. Zijderveld[1], S. Abghari[1], S. van Heukelum[1], P. Tiesinga[2] and J.C. Glennon[1]

[1]Department of Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen, The Netherlands.
[2]Department of Neuroinformatics, Radboud University, Nijmegen, The Netherlands. jglennon7@gmail.com

Abstract

Due to their scope for (opto)genetic intervention, mice have become a dominant model for studying visual processing in a behavioural context (Carandini & Churchland, 2013; Bennett et al., 2013). While visual tasks for mice have been drastically expanded in recent years (Glickfeld et al., 2013; Poort et al., 2015], most still face at least one of two issues: Long training times (see Figure 1), and binary (hit/miss), roughly timed behavioural read-outs, requiring broad averaging across trials. This presents a bottleneck in relating neuronal dynamics to cognitive processing: The scope and precision with which neuronal activity can be measured and manipulated have improved drastically, yet potential insights are limited if we cannot relate neuronal activity to equally precise behavioural measures, and if measurements are made mainly in over-trained animals. These are key limitations of current operant behavioral training setups.

Here we present a visual discrimination paradigm for head-fixed mice that relies on 'foraging-like' navigation in a virtual environment (Schmidt-Hieber & Hausser, 2014) that has been validated in mice with C57Bl/6 or BALB/cJ genetic backgrounds. All experiments were approved by the ethical review board of Radboud University and conducted according to Dutch ethical committee guidelines for use of animals in experimentation under licence number RU-2013-238. In all cases, animals underwent surgery for the implementation of a head plate to allow head fixation which was followed by a one week recovery period. This was carried out under isoflurane anaesthesia with appropriate administration of analgesics. C57Bl/6 mice were chosen as they are the dominant mouse genus in use for behavioural and genetic experiments. The task requires animals to discriminate between sinusoidal gratings at orientation differences of 90° down to 5°. Thus, the mice are trained to respond to horizontal line cues whereby the most horizontal in a choice between two cues is rewarded. The lines vary in angle somewhere between 0° and 90° with 5° increments (see Figure 2). The associated training scheme allows animals to learn precise visual discrimination in four training sessions or less (two sessions or less of training in visual discrimination), corresponding to 358±55 trials (corresponding to 11-12 sessions in conventional operant chambers) counting from the first contact with the experimental setup. It allows for precise single-trial read-out of behavioural choices, including reaction time, response accuracy and response confidence. The mechanism underlying reduction in training times is unknown. However, it could be speculated that it is related to the 'active' acquisition of the task during running rather than the normal 'passive' acquisition where the animal is still or able to ignore the task in a normal operant box that is the difference. It is well established that exercise increases neurogenesis in areas such as the dentate gyrus. This may give some additional capacity towards task learning in those learning a task while active.

Variations in task design are easily implemented to test for visual discrimination between various stimuli and visual detection. Furthermore, advanced processing elements such as attentional cues and reward rules are enabled in the virtual environment allowing the creation of operant behavioural tasks such as reversal learning. This makes it possible to relate neuronal responses (measured by tetrode recording) and manipulations (such as optogenetics) in regions such as mouse visual cortex V1 to subsequent behavioural choices on a trial-by-trial basis. Furthermore, the virtual environment enables the capitalisation of features such as fast training times to track changes in stimulus processing and behavioural strategies. This has four key advantages:

4. Coupling of neuronal activity and optogenetic manipulation more precisely to behavioural changes in the temporal domain.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

5. Rapid training of animals such that young animals can be trained and tested on an operant task in the juvenile life phase which is key for the study of neurodevelopmental disorders such as attention deficit hyperactivity disorder and autism. Currently, existing operant behavioural training setups take so long to train animals on a task (6-12 weeks) that they are tested for task performance during the adult and not the juvenile life phase. This allows improvement matching in terms of testing the early life phase which is relevant to neurodevelopmental disorders such as autism.

6. Reduction in the overtraining of animals reduces the contribution of memory processing to task performance thereby enabling a more naturalistic study of task performance. (While the virtual environment is programmable in terms of tasks presented and these are artificial tasks rather than real-world tasks the animals would normally face (foraging; food-seeking; predator avoidance, etc.); we intended that at least the virtual environment was more natural in allowing free running within an environment (albeit virtual) than conventional operant behavior).

7. Accelerated training time of animals for operant tasks increases the throughput of animals able to be tested which is important in drug discovery during the biological lead optimisation and clinical candidate selection phases. Drug testing can therefore be performed within one to two months rather than four to five months as with conventional operant training paradigms.

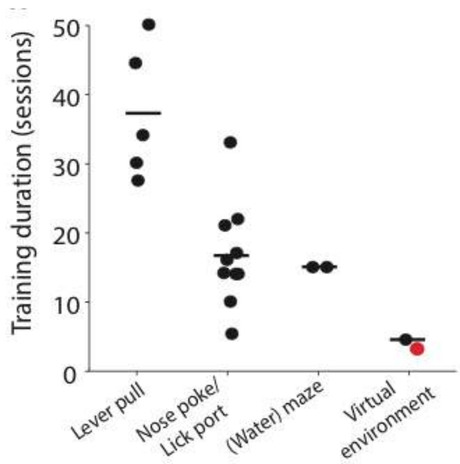


Figure 1. Comparison of training duration (number of sessions) in virtual environment with other conventional operant task paradigms.

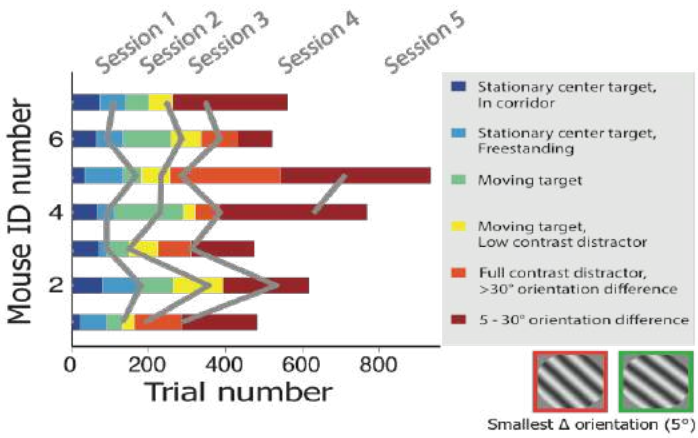


Figure 2. Schematic of training setup and experimental design in virtual environment.

References

1. Carandini, M., Churchland, A.K. (2013). Probing perceptual decisions in rodents. *Nature Neuroscience*, **16**(7), 824-831.
2. Bennett, C., Arroyo, S., Hestrin, S. (2013). Subthreshold Mechanisms Underlying State-Dependent Modulation of Visual Responses. *Neuron* **80**(2), 350-357.
3. Glickfeld, L.L., Histed, M.H., Maunsell, J.H. (2013). Mouse primary visual cortex is used to detect both orientation and contrast changes. *J. Neurosci.* **33**(50),19416-22.
4. Poort, J., Khan, A.G., Pachitariu, M., ... Keller, G.B., Mrsic-Flogel, T.D., Hofer, S.B. (2015). Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. *Neuron* **86**(6), 1478-1490.
5. Schmidt-Hieber, C., Häusser, M. (2014). How to build a grid cell. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369**(1635), 20120520.

Inference of Cattle Behavior on Rangeland Using Animal-Borne Sensors

Y. Nevo[1], A. Arieli[2], S. Navon[1], H. Baram[1], A. Belalu[1], E. Ben Moshe[1] and E.D. Ungar[1]

1 Department of Natural Resources, Institute of Plant Sciences, Agricultural Research Organization – the Volcani Center, Bet Dagan, Israel. eugene@volcani.agri.gov.il

2 Department of Animal Sciences, Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University Jerusalem, Rehovot, Israel.

Introduction

Rangelands grazed by domesticated livestock constitute about one quarter of the terrestrial land area of the planet. Rational use of these areas can improve outputs of secondary (animal) production and also contribute to the sustainability of their plant and animal life. The influence of the producer on production efficiency and on the environment depends on decisions that he takes on a daily basis and in long-term planning. The options open to the producer are management decisions that he implements routinely: when to introduce animals into a paddock and when to remove them; animal density; and whether to provide supplementary feeding. Such decision making is based currently on familiarity of the producer with the rangeland and experience. Continuous and reliable information on the state of the rangeland and the cattle is likely to engender more rational decision making. The monitoring of animal behavior and the discernment of changes that occur in it could be an indicator as to the quantity and quality of vegetation and could constitute valuable information regarding the physiological and even health status of the animals. There have been studies that dealt with the relationship between the information received from various monitoring devices and behavior [1, 2, 3, 4, 5], but the most suitable type of sensor and the level of precision that can be attained remain open questions.

The objective of this study was to evaluate our ability to infer the behavior of beef cattle on the basis of data provided by three different sensors, separately and together: GPS, pedometry and acoustic monitoring. The combination of sensors was intended to characterize the different factors through which it would be possible to infer cow activity. Pedometry enables quantification of cow activity related to leg movements and body position (upright versus lying). GPS provides cow position, from which variables based on changes in position can be calculated. Acoustic monitoring enables jaw movements to be monitored.

Materials and Methods

Experiments were approved by the animal experimentation ethics committee of the Agricultural Research Organization (approval IL 385/12). The pedometer used was the IceTag (IceRobotics, Edinburgh, Scotland), which is the most sophisticated pedometer that is commercially available. Because of the high cost of commercially available GPS collars for animals, we used a standard GPS device (i-gotU GT600, Mobile Action, Taiwan) for which we assembled an external battery and a home-made collar. For acoustic monitoring we used an audio recording device (Sansa clip+, SanDisk, USA) for which we assembled a specialized external microphone and an external battery. The equipment records sound vibrations as transmitted by jaw movements to the horn of the cow. The signal processing was performed by a dedicated software package that generates the timeline of jaw movements [6] and distinguishes between bouts of intake and bouts of rumination [7]. For the first time, this study implemented acoustic monitoring for time periods greater than a few hours, and in one of the deployments we achieved a recording duration of eight days.

The study was conducted in rangeland paddocks belonging to Kibbutz Ein Hashofet, Israel (see Picture 1). Twelve mature, local cows of mixed breeds participated in the first round of measurement (summer 2012), and a further 12 cows participated in measurement rounds 2 (winter 2013), 3 (spring 2013) and 4 (summer 2013). During the measurement rounds, video clips were filmed in order to enable detailed and precise description of behavior. Each clip focused on one cow and lasted five minutes, and in total there were 296 clips.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

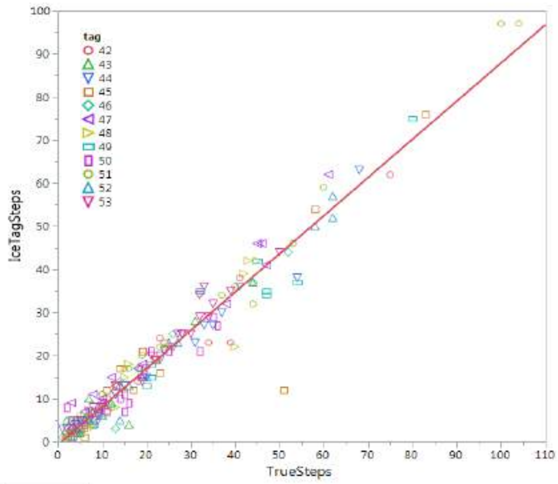
Picture 1. Cow wearing pedometer (rear, left leg), GPS collar and acoustic sensor (black box on left horn) in a paddock of herbaceous rangeland vegetation at Kibbutz Ein Hashofet, Israel.

Data analysis included three main phases: 1) data validation of the outputs received from each device; 2) calibration of device data in order to find the relationship between the measured variables and the behavior performed; 3) computation of the daily behavioral pattern by application of the calibration equations to the entire database. The first two phases were based on a comparison of sensor data to that coded from the video clips. The clips were coded twice, once to generate the timeline of steps, and a second time to generate the detailed activity timeline which included the following categories of activity: active grazing; intermittent grazing; traveling (without grazing); lying down; standing still; loitering; undefined; and unknown. An algorithm was applied to the combination of these two timelines to determine a single characteristic behavior for each clip. The behaviors were classified as graze, rest, travel (GRT) or as graze, stand, lie, travel (GSLT). The ability to predict behavior from sensor data was evaluated using partition analysis. Using the calibration equations derived by partition analysis, the most likely behavior was computed for each 5-minute period of the day. This was summarized at the half-hourly level and the allocation of time among the different behaviors was computed.

Results

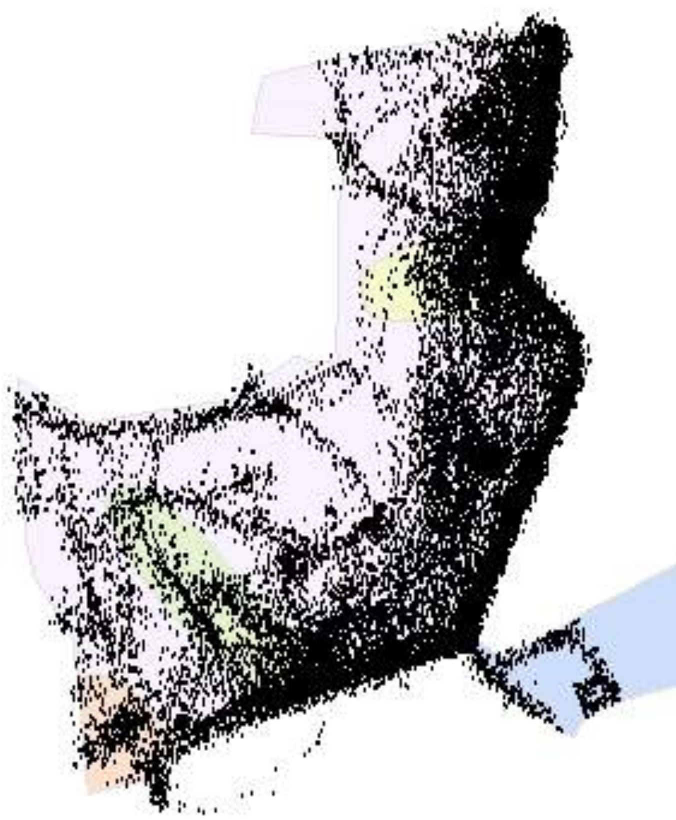
The validation phase showed excellent results for the three devices. The IceTag device was found to be accurate in overall step count relative to observed counts (R^2 = 0.97; see Figure 1). Nevertheless, a synchronized comparison of the two timelines revealed a matching level of 60% when a tolerance of 0.5 seconds was allowed.

Figure 1. Correspondence between the number of steps coded from video analysis (X axis) and the number of steps registered by the IceTag pedometer (Y axis). Symbols correspond to individual cows. Line shows linear fit.



In an examination of the GPS data, 94% of locations fell within the paddock fence line, and this increased to 99.7% when a 5-meter buffer beyond the fence line was included (see Figure 2).

Figure 2. Pattern of paddock utilization by a group of 12 cows over one month in January 2013 based on i-gotU (Mobile Action, Taiwan) GPS device operating with 10 s sampling interval.



The acoustic monitoring system showed good correspondence between the rate of jaw movements during five minutes to the time spent grazing during the same 5 minutes ($R^2 = 0.82$; see Figure 3).

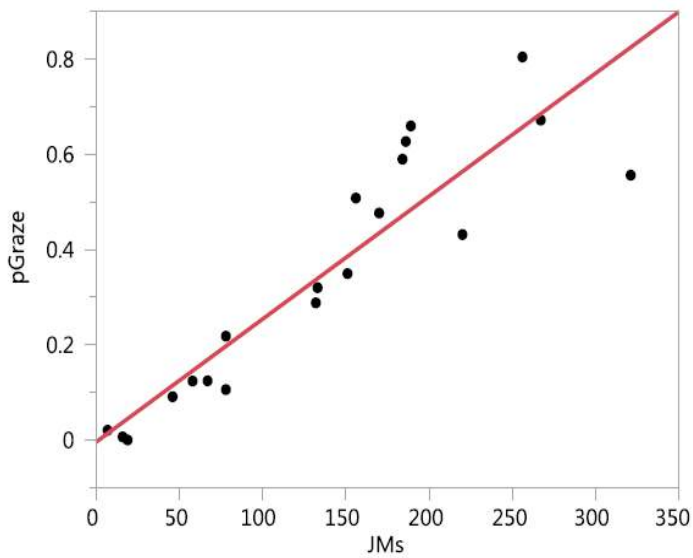


Figure 3. The relationship between the total number of jaw movements (JMs) performed during a 5-minute period as determined by acoustic monitoring (X-axis) and the proportion of time allocated to active grazing (pGraze) during the same 5-minute period as determined by analysis of video recording (Y-axis).

The calibration equations for the prediction of behavior according to the GSLT classification, and on the basis of GPS and pedometer variables together, correctly classified 89% of observationss (see Table 1 analysis 7). The use of variables from just one device yielded lower values: 85% for classification based on the IceTag device (see Table 1 analysis 5), and 69% for classification based on GPS (see Table 1 analysis 6). Acoustics is not relevant to the separation of standing and lying but is was effective in separating grazing and standing. For GRT classification of behavior, GPS and pedometer variables together correctly classified 89% of observations (see Table 1 analysis 3), and, on the basis of one device only, the rates were 89% for GPS (see Table 1 analysis 2) and 87% for IceTag (see Table 1 analysis 1). Combining acoustic data did not improve the rate of correct classification (86%; see Table 1 analysis 4) but this result was influenced apparently by a low observation sample size. Calculation of expected grazing time according to acoustic variables only was carried out using the regression equation between grazing time extracted from the observations and rate of jaw movements (see Figure 3). We found that the separation of lying and traveling was not difficult and the main problem remained the separation of standing and grazing. In this respect the rate of correct classification reached 70% when the analysis was based on IceTag and GPS variables.

Table 1. Confusion matrices showing the level of correct classification of activity based on partition analysis of data from measurement round 4. Values for observed and predicted activities are number of observations; n is total observations. Upper and lower sections of table are for GRT (Graze, Rest, Travel) and GSLT (Graze, Stand, Lie, Travel) classification of activity, respectively. Column Correct shows the percentage of observations that were correctly classified.

			Predicted activity					
Analysis	Device	Observed activity	Graze	Rest	Travel		N	Correct (%)
1	IceTag	Graze	4	15	0		19	21
		Rest	1	105	1		107	98
		Travel	0	0	4		4	100
		Overall					130	87
2	GPS	Graze	19	0	0		19	100
		Rest	13	93	1		107	87
		Travel	0	0	4		4	100
		Overall					130	89
3	GPS + IceTag	Graze	19	0	0		19	100
		Rest	13	93	1		107	87
		Travel	0	0	4		4	100
		Overall					130	89
4	GPS + IceTag + acoustics	Graze	8	0	0		8	100
		Rest	1	7	2		10	70
		Travel	0	0	3		3	100
		Overall					21	86
			Graze	Lie	Stand	Travel		
5	IceTag	Graze	0	0	19	0	19	0
		Lie	0	67	0	0	67	100
		Stand	0	0	39	1	40	98
		Travel	0	0	0	4	4	100
		Overall					130	85
6	GPS	Graze	19	0	0	0	19	100
		Lie	0	67	0	0	67	100
		Stand	13	26	0	1	40	0
		Travel	0	0	0	4	4	100
		Overall					130	69
7	GPS + IceTag	Graze	19	0	0	0	19	100
		Lie	0	67	0	0	67	100
		Stand	13	0	26	1	40	65
		Travel	0	0	0	4	4	100
		Overall					130	89

Application of the calibration equations derived from IceTag and GPS variables only, to the entire dataset, yielded suspect allocation of behaviors, with a bias towards too much grazing time. A correction to this equation based on acoustic data yielded a behavioral allocation more in line with the expected results. These results were similar to the calculation of grazing time derived from acoustic monitoring only. The daily summary of behavioral allocation for the spring and second summer measurement periods yielded the following results: walking – 0.4 and 0.4 hours/day, respectively; standing – 5.9 and 8.0 hours/day, respectively; lying down – 10.3 and 9.4 hours/day, respectively; and grazing – 5.5 and 5.8 hours/day, respectively. These values are consistent with values published in the scientific literature.

Conclusions

By far the most important separation of behaviors in a management context is that between grazing and non-grazing. Neither GPS nor pedometry can form the basis for particularly good prediction models, each for its own reason. Animal-borne GPS devices do not yet provide the precision and sensitivity required to distinguish definitively between a (more-or-less) stationary and a slowly-moving animal on a 5-minute time scale. The problem is with the technology. Pedometry is able to generate an accurate timeline of leg activity. It is the animal that is not totally still when standing, and moves very slowly when grazing, that complicates the interpretation. There is a highly variable pattern of stepping within observations classified as grazing. Acoustic monitoring taps into the defining process of active grazing. If the signal processing is reliable, then defining periods which are neither grazing nor rumination is straightforward. Grazing and rumination are easily distinguished on the basis of the pattern of jaw movements, with rumination having a readily identifiable pattern. One of the most important results of the study was demonstration of the ability to implement acoustic monitoring under commercial field conditions for a significant time period. Analysis of behavior which included data from acoustic monitoring indicated that use of this data enables grazing and standing to be separated more effectively, compared to the other sensors that were examined. But further work is needed in order to verify this claim more solidly. The combination of data from different devices, and especially from acoustic monitoring, together with location data should enable many questions related to rangeland utilization by grazing animals to be addressed.

References

1. Turner, L.W., Udal, M.C., Larson, B.T., Shearer, S.A. (2000) Monitoring cattle behavior and pasture use with GPS and GIS. *Canadian Journal of Animal Science* **80**, 405–413.
2. Moreau, M., Siebert, S., Buerkert, A., Schlecht, E. (2009) Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour. *Applied Animal Behaviour Science* **119**, 158–170.
3. Umstätter, C., Waterhouse, A., Holland, J.P. (2008) An automated sensor-based method of simple behavioural classification of sheep in extensive systems. *Computers and Electronics in Agriculture* **64**, 19–26.
4. Ungar, E.D., Henkin, Z., Gutman, M., Dolev, A., Genizi, A., Ganskopp, D. (2005) Inference of animal activity from GPS collar data on free-ranging cattle. *Rangeland Ecology and Management* **58**, 256–266.
5. Ungar, E.D., Schoenbaum, I., Henkin, Z., Dolev, A., Yehuda, Y., Brosh, A. (2011) Inference of the activity timeline of cattle foraging on a Mediterranean woodland using GPS and pedometry. *Sensors* **11**, 362–383.
6. Navon, S., Mizrach, A., Hetzroni, A., Ungar, E.D. (2013) Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring. *Biosystems Engineering* **114**, 474–483.
7. Ungar, E.D. (2014) Acoustic monitoring to infer activity and intake of grazing animals. *Abstract Book of International Conference on Ecoacoustics* (Paris, 16–18 June 2014), 78.

Monitoring Chickens' Position in a Free-range Area Using an Ultra-wideband System

L.M. Stadig[1,2], B. Ampe[1], J. Criel[3], T. Bas Rodenburg[4], B. Reubens[5], F.A.M. Tuyttens[1,2]

[1]Animal Sciences Unit, Institute for Agricultural and Fisheries Research, Melle, Belgium. lisanne.stadig@ilvo.vlaanderen.be
[2]Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium.
[3]Sensolus, Ghent, Belgium.
[4]Behavioural Ecology Group, Wageningen University, Wageningen, the Netherlands.
[5]Plant Sciences Unit, Institute for Agricultural and Fisheries Research, Merelbeke, Belgium

Introduction

Chickens with free-range access often make limited use of the outdoor area. A more intensive use and a better distribution over the available area could have several advantages regarding animal welfare and soil and water quality. To this end, we are testing several strategies to improve chickens' free-range use, such as different shelter types and enrichment early in life. To assess the efficacy of these strategies, free-range use is until now mainly monitored using visual observations. However, these observations have several disadvantages. Firstly, the presence of an observer can influence the animals' behaviour. Secondly, the accuracy of human observations may not be optimal, and the possibility of observer bias can never be ruled out. Thirdly, observations are very time-consuming, and can rarely be done at the same frequency as automatic monitoring. Lastly, it is very difficult to track individual birds with visual observations. In addition, in our studies we use short rotation coppice (SRC) with willow as a shelter type, which is a very dense vegetation, making visual observations even more complicated. Therefore, our aim is to develop an effective and efficient system which can automatically register the positions of the chickens with access to a free-range area.

Materials and Methods

To develop a system which was suitable for our research, we had to overcome several challenges. Firstly, because chickens often stay close to their house, the accuracy of the system had to be high, preferably <50 cm. This is necessary to be able to detect differences e.g. in the mean distance from the house between the two shelter types, or whether a chicken is inside or outside the house. To meet this demand, we chose to use Ultra-Wideband (UWB) technology, which is a radio technology capable of transmitting large amounts of data over short range at low power. Such a system allows calculating the position of sensors with high precision (<50 cm), based on time-of-flight. The system we used consists of custom sensors, which are fitted on the chickens, trackers on fixed positions on the field (Figure 1) to measure the positions of the sensors, a gateway that aggregates information from the trackers and sends it to the cloud, and cloud software.

The second challenge was that the sensors had to be small and light enough for chickens to carry. The final version of the sensor weighed 36 grams, and its dimensions were 75x49x17 mm. The sensor was placed in a custom-made backpack with elastic, adjustable bands, which were placed around the wings to secure the sensor on the chicken. Even though the sensors were kept as light as possible, wearing it might affect the chickens' behaviour. They might e.g. walk less when wearing a sensor compared to without. Therefore, we are currently performing an experiment to quantify this possible effect, by video recording chickens with and without sensors from 5-10 weeks of age and then comparing their behaviour.

Third, battery life had to be at least 35 days with a sample rate of once every 2 minutes. This proved difficult, given the minimisation of sensor size and weight. So far, we have performed pilot tests in two production rounds of slow-growing broiler chickens (Sasso hybrid), in which we tested lifetime and signal strength of the sensors. The chickens were given outdoor access from day 35 to 70 (slaughter age). In round 1, 30 chickens were equipped with sensors from day 46 to 58, and in round 2 20 chickens had sensors from day 37 to 66. The first round, the sample rate was set at 20 sec, and the sensors lasted for a maximum of 13 days. For the second round, adjustments were made to the firmware of the sensor. At a sample rate of once per minute, maximum life time was now 30 days. To further improve battery life, the sensors are currently being altered by making changes to the embedded battery optimization algorithms. For the final version we aim at a battery life of 35 days (based on a sample rate of once every 2 minutes).

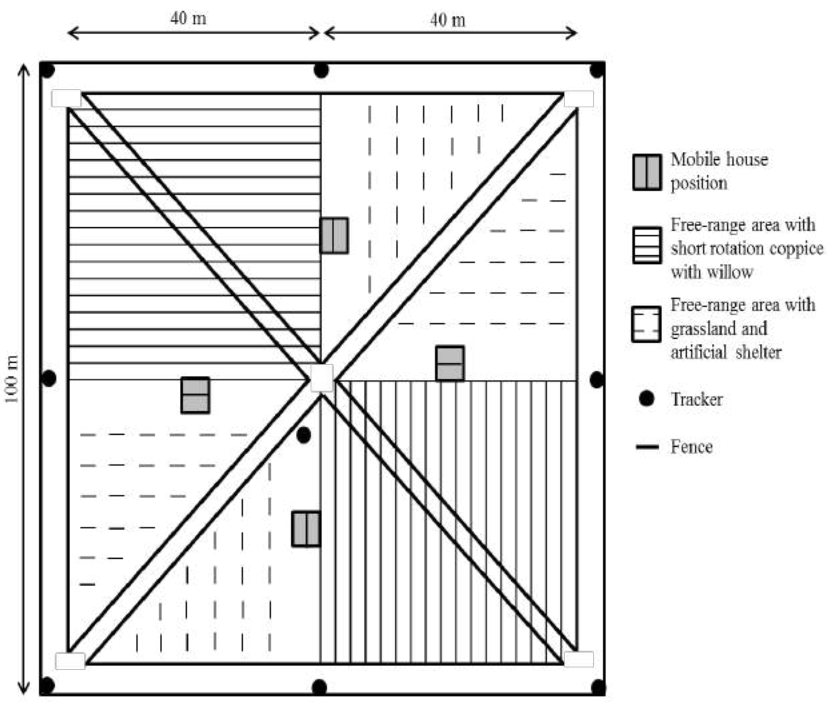


Figure 1 - Plan of the experimental field.

Fourth, the sensors' signal strength had to be strong enough to send the signal through the SRC. On the grassland there were no problems, but the preliminary tests showed that coverage in the SRC was insufficient. Therefore, adjustments were made to the set-up of the trackers. With the final version of the sensors, the furthest 5% of the SRC field could not be covered. Therefore, for the coming experiments, the experimental field will be decreased in size to have coverage in all areas. This is not expected to have a large effect on free-range use as most chickens (95th percentile) stay within 5.4 and 4.5 m of their houses, for SRC and grassland with artificial shelter respectively.

During the two production rounds, it was also assessed whether the sensors would have a negative effect on production parameters or animal welfare measures. To this end, weight at the end of production round 3 was recorded, and hock and foot pad dermatitis, gait and cleanliness were scored in an equal number of birds with and without sensors (same distribution of male:female). All animal experiments were approved by the Ethical Committee of the Institute for Agricultural and Fisheries Research.

Results

Preliminary analysis showed a minimal number (0.1%) data points at which a chicken was assigned a position in a field it could not be in. Another 7.2% of the night data points wrongly assigned chickens to be outside of their houses. The cause of this (e.g. incorrect house coordinates or inaccurate sensor positions) needs to be further investigated. The data showed that chickens spent most time in their house (71.6% of daytime), when they were outside they were more often in the willows than in the grassland 18.4 vs. 9.9% of daytime). The average distance from the house was higher when chickens were in the willows as compared to the grassland (2.19 vs. 1.17 m). Welfare indicators (foot pad dermatitis, hock dermatitis, gait, cleanliness) and slaughter weight were not negatively affected by wearing the sensors.

Conclusion

There were several challenges that had to be overcome in designing this system for automatic monitoring of chickens' positions, and the lifetime and accuracy of the sensors remain a critical point. However, the first test results were promising, and after the final alterations have been made, the expectation is that the system will be suitable to track chickens in an outdoor setup.

Automated Stress Detection in Police Horses

D. Piette, V. Exadaktylos and D. Berckmans

Department of Biosystems, KU Leuven
{Belgium. deborah.piette}@biw.kuleuven.be

Introduction

When it comes to equestrian disciplines, welfare and safety are amongst the most discussed topics. Recently the emergence of equitation science and the increasing awareness of animal welfare have resulted in more attention for horse welfare and for horse and rider safety in the equine field. An important share of the research in equitation science focuses on the assessment and improvement of welfare and safety and on identifying the factors they can be influenced by. One of the most studied factors that affects the welfare and safety of the horse and its rider is psychological stress or mental state. Not only does stress have an impact on welfare and safety, it is also known to have an effect on health and performance. This effect can be either positive or negative. For instance small amounts of stress can sharpen the senses and improve performance (Westman & Eden, 1996), whereas stress levels that are too high can lower performance (Peeters, Closson, Beckers, & Vandenheede, 2013). Stress with a positive effect is referred to as eustress, in contrast to distress which has a negative effect. Stress that has a negative effect on the welfare, safety, health and performance of horses and their riders must be avoided at all times. Monitoring the mental state of horses and riders in real-time can help to detect distress in order to address it, but can also be used to prevent distress from occurring. By doing so, the welfare, safety, health and performance of horses and riders can be improved.

At the mounted police in Brussels, Belgium, ensuring the welfare, safety, health and performance of the riders and the horses is important to be able to provide an optimal service to society. In that regard, proper training and careful selection of candidate riders and horses are key. The selection of candidate horses in particular, is one of the main concerns of the mounted police today as police horses have to be able to remain calm in very challenging situations. For this purpose the behaviour of candidate police horses and their response to stressful events must be thoroughly analysed before purchase. To date the decision whether or not to buy a horse is mainly a subjective one. It is largely based on the previous experience of the mounted police trainers and riders, and on their gut feeling. However, the high subjectivity of this decision does not mean that it is a bad decision. As a matter of fact at the mounted police in Brussels the candidate horses that are evaluated as suitable and are subsequently purchased, mostly live up to their expectations. However occasionally a horse is bought that does not evolve into a good police horse, which means loss of time and investment. This is where an additional objective measure to determine the suitability of candidate police horses might prove useful to further increase the number of good decisions and to reduce the costs of the mounted police associated with the selection and training of horses.

Objectives

The first aim of this work is to assess the behaviour of police horses in a continuous and automated way by monitoring their mental state using wearable technology. The second goal is to find an objective measure based on the mental state of the horses that can predict the future performance of candidate police horses to help the mounted police in their decision on whether or not to buy a horse.

Study design and subjects

Four protocols were composed in consultation with the mounted police to evaluate the reaction of police horses to conventional mounted police tasks: a 'going to work' test (GTW), an obstacle test (OT), a flexibility test (FT) and a 'coming from work' test (CFW) (Piette, 2015). The GTW serves to observe the reaction of the horse when

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

it is prepared to go to work and when it is brought to the training arena. In the OT the horse has to pass a series of obstacles: walk over plastic bottles, walk through a curtain with plastic straps, stand still while the rider knocks over a barrel and stand still next to a speaker playing a recording of gun sounds, shouting and clapping sounds. In the FT the reaction of the horse to a change in rider is examined. The CFW serves to observe the horse's reaction when it comes back from the training arena and is prepared to go back into its box.

For the experiments, 9 horses (3 mares and 6 geldings, aged 4-15 years) were divided into four categories by mounted police riders, based on their experience level and suitability as a mounted police horse: good beginner, bad beginner, good experienced and bad experienced. Beginner horses are horses that have recently entered the training period at the mounted police. Experienced horses, in contrast, have been operational for several months up to several years. The label 'good' or 'bad' has been given to the horses previous to the start of the experiments by the mounted police riders based on their subjective experience with the horses and based on the progress the horses have been making during their training period or operational career. Each category contained two horses, with the exception of the category 'good beginner' which contained three horses. The horses were ridden by 9 riders (4 female and 5 male, aged 24-50 years) of the mounted police. All measurements were performed twice with all horses but one good beginner horse, making sure that no horse was ridden twice by the same rider to take into account the effect of the rider on the horse. During the experiments the heart rate and physical activity of the horses were measured at a sampling frequency of 1Hz using a Polar heart rate belt and a Zephyr module respectively. The heart rate was measured in beats per minute (bpm), the physical activity was measured in Vector Magnitude Units (VMU). All experiments were recorded on video. In total 68 measurements were carried out with 9 horses, which gave a total of 15 hours of data. The experiments were approved by the Animal Ethics Committee of KU Leuven. Data processing was performed in Matlab (Matlab).

Methodology

The heart rate of a living organism can roughly be divided in three parts. A first part is related to the organism's metabolism, a second part is correlated with the physical activity of the organism and a third part links to the organism's mental state or stress. An illustration of how heart rate can change as a function of metabolism, physical activity and mental state is shown in **Figure 1**.

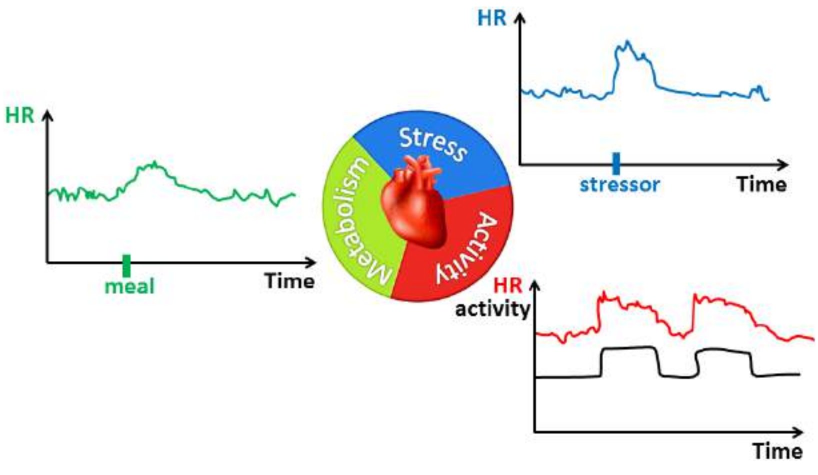


Figure 1 Illustration of the three most important determinants of heart rate (GH, 2014).

The scope of this work is to estimate the mental state of police horses. In order to estimate the mental component of the heart rate, the components 'physical activity' and 'metabolism' have to be quantified. The physical activity is estimated by measuring the activity of the horses. The metabolism of the horses is assumed constant during the short experiments and is consequently not further taken into account. (Jansen)

Using the measured physical activity of the horses, a simulated heart rate was calculated that only reflects changes in heart rate resulting from changes in physical activity of the horses. This was done by means of mathematical modelling. Since a change in the organism's activity induces a change in its heart rate, the organism can be seen as a Single-Input-Single-Output (SISO) system with activity as the input variable and heart rate as the output variable (Jansen). The mathematical relation between input and output variable can then be described by a linear AutoRegressive model with eXogenous input (ARX).

The general equation of a linear ARX model is given below:

$$y(t) + a_1 y(t-1) + \ldots + a_n y(t-n) = b_0 u(t) + b_1 u(t-1) + \ldots + b_m y(t-m)$$

with:

with:	y:	output variable
	u:	input variable
	t:	time instant of the measurement
	$a_1, \ldots, a_n$:	a-parameters
	$b_0, b_1, \ldots, b_m$:	b-parameters

Note that the output variable $y(t)$ is the simulated, or modelled, heart rate of the organism and that the input variable $u(t)$ is the measured activity of the organism.

In order to determine the model order of the ARX model, model training must be performed. Model training consists of calculating the values of the a- and b-parameters that result in the modelled heart rate $y(t)$ that approaches the measured heart as close as possible, and this for different model orders. A model with n_a a-parameters, n_b b-parameters and a time delay τ_d, is said to have model order $[n_a, n_b, \tau_d]$. For this work, n_a and n_b were varied between one and three to avoid models that are too complex. The time delay τ_d was varied between one and five seconds.

Since the objective is to compute a heart rate that solely contains the heart rate component related to the organism's activity, the model training must be carried out on a part of the measurement during which the organism is in a neutral mental state. Equally important is that the training set contains a step-up, i.e. a transition between two steady states. Incorporating a step-up in the training data ensures that all frequencies of the system have been excited. Figure 2: illustrates the selection of a training set for a good experienced horse during the FT. Since the chosen training set is located before the stressor, which is the change in rider, it can be assumed that the horse is in a neutral mental state. Additionally, a step-up was incorporated in the training set, namely the transition from standing still to walking (step 1 to step 2).

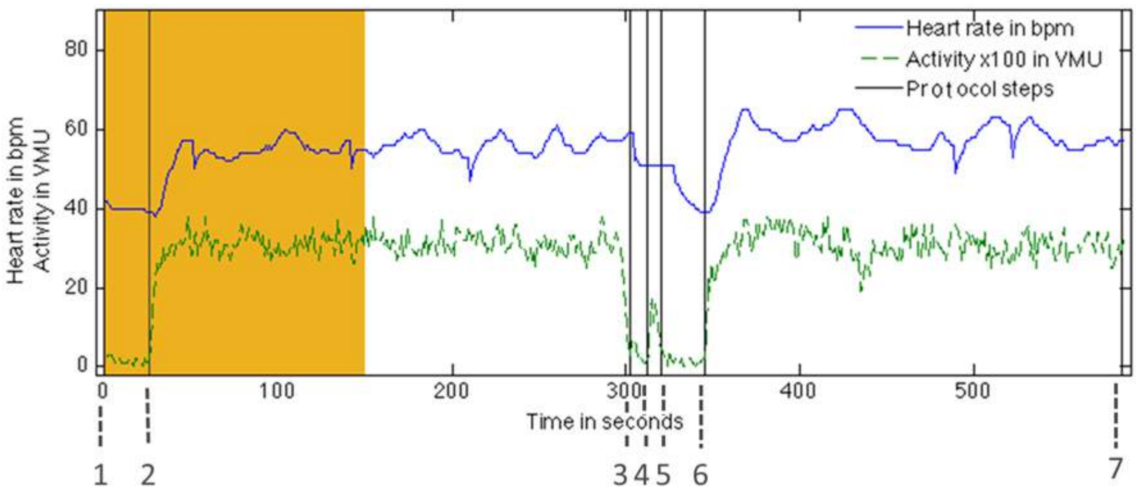


Figure 2: Choice of model training set for a good experienced horse during the FT. The different steps in the FT protocol indicated in the figure are: 1) Standing still, 2) Walking, 3) Rider 1 dismounts, 4) Rider 2 mounts, 5) Standing still, 6) Walking, 7) End. The training set selection is marked as a coloured block in the figure.

The comparison of the modelled and measured heart rate for the different model orders was done visually and by means of the Young's Identification Criterion (YIC) (Young, Recursive estimation, forecasting and adaptive control, 1989; Young, Recursive Estimation and Time-Series Analysis, 2011). During visual analysis attention was paid to the trend of the modelled heart rate which must be as close as possible to the trend of the measured heart rate. The YIC-value should be minimised (large negative value) as much as possible to ensure model stability. **Figure 3** shows the measured heart rate and modelled heart rate for a model of order [2, 2, 1] applied to the training data shown in **Figure 3**.

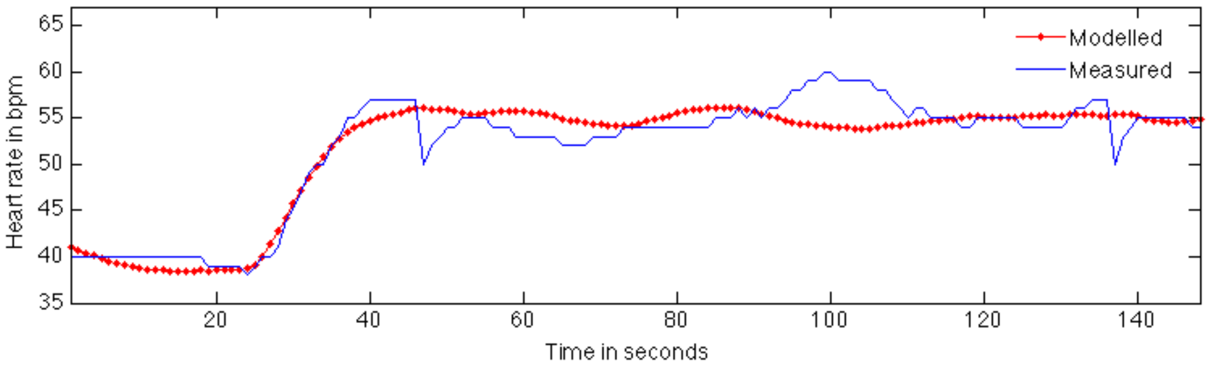


Figure 3: Visual comparison of modelled and measured heart rate for a good experienced horse during the FT.

As can be observed in Figure 3, the modelled heart rate follows the general trend of the measured heart rate well. Especially the transition zone between the two steady states is modelled properly. For these reasons the model order [2, 2, 1] is considered suited for this measurement. In fact, this model order was found to be appropriate for most of the measurements. Hence, it was decided to work with fixed model order [2, 2, 1] to obtain the modelled heart rate and the associated a- and b-parameter values for all measurements.

Once the a- and b-parameters are obtained from the model training, the linear ARX equation can be applied to the entire measurement, resulting in a modelled heart rate describing the organism's heart rate component that is solely related to activity. The positive difference between the measured heart rate and the modelled heart rate is a measure for the organism's psychological mental state. By taking the window average of this difference using a window of 20 seconds, a smoother stress graph is obtained. Finally, a normalisation of this graph is carried out to obtain the organism's relative stress in percentage during the measurement. This is illustrated in Figure 4 for a bad experienced horse during the OT, which clearly shows that stress is detected when the horse is subjected to the stressor (the course of obstacles).

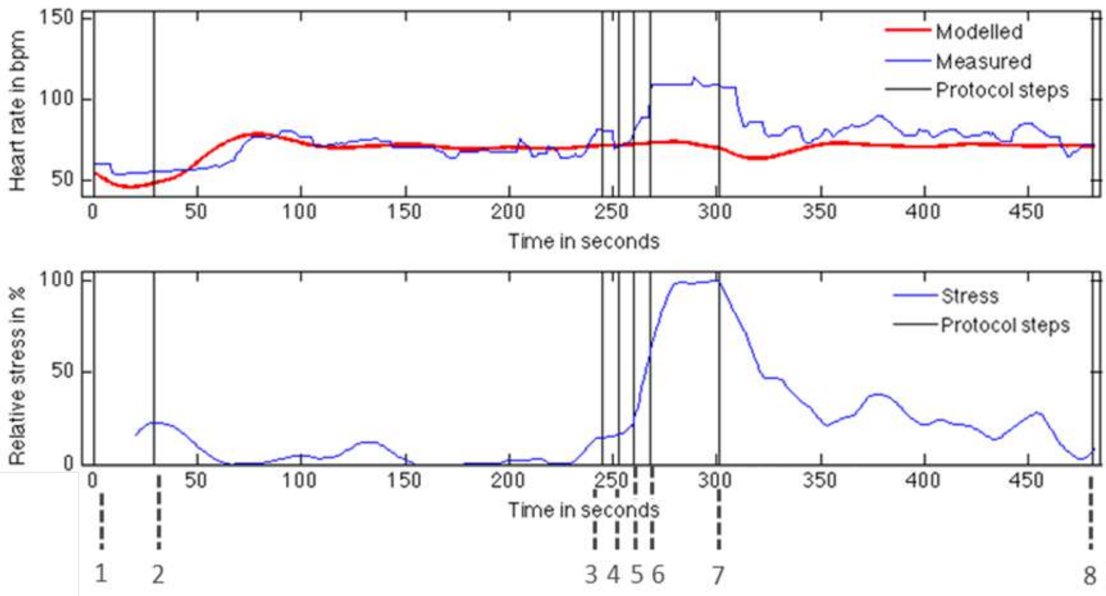


Figure 4: Obtaining the relative stress graph for a bad experienced horse during the OT. The different steps in the OT protocol indicated in the figure are: 1) Standing still, 2) Walking, 3) Plastic bottles, 4) Curtain with plastic straps, 5) Rider knocks over a barrel, 6) Sound being played, 7) Walking, 8) End.

For the validation of the mental state detection of the horse, the relative stress graph must be validated against the gold standard. For this work, behaviour was chosen as the gold standard for stress in horses and was obtained using an equine ethogram. Since the focus of this work lies in detecting negative mental state, a list of thirteen types of negative equine behaviour chosen from the most frequently used ones in literature has been composed (ears focused, ears flat, tail beating, tail high, head high, head tossing, bucking, rearing, hitting, jumping, refusing, fleeing, scraping). Evaluation of the horse behaviour was done by one person with experience in horse behaviour in order to ensure consistency. Using the video recordings of the horses together with a custom labelling tool designed in Matlab (Matlab), the occurrence of each type of behaviour was stored for every second of the experiment. In other words, for every second of the measurement the amount of different types of negative behaviour shown by the horse was stored. This resulted in continuous equine behaviour data with the same 'sampling rate' as the measured heart rate and physical activity, namely 1Hz, which enables validation of the model used for stress detection. The same window averaging and normalisation procedure as for the modelled heart rate was applied to the labelled behaviour to obtain relative negative behaviour in percentage. An illustration of the labelled negative behaviour and the relative negative behaviour derived from it can be found in **Figure 5**.

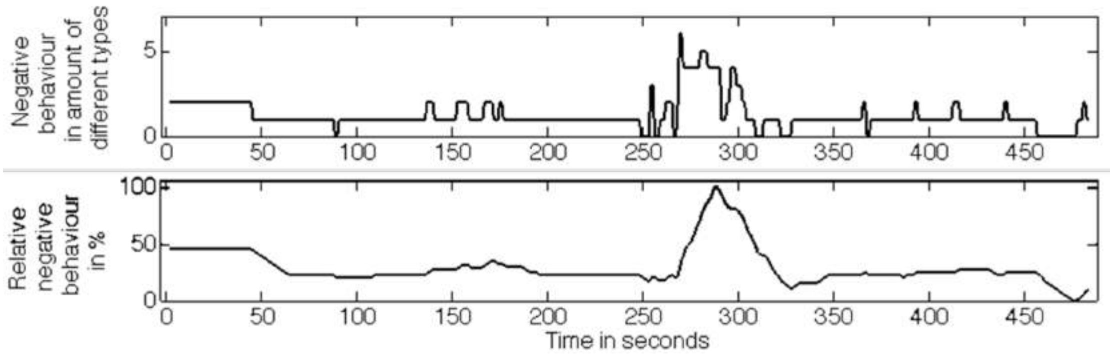


Figure 5: Obtain the relative negative behaviour in percentage from the labelled negative behaviour for a bad experienced horse during the OT.

For the validation of the stress detection of the horse, the relative stress was compared with the relative negative behaviour. First, a peak detection algorithm was applied to both time series (minimal peak distance = 50 seconds, minimal peak height = 10%). Positive peaks were matched between relative stress and relative negative behaviour in a range of ± 50 seconds. An example is shown in Figure 6. Applying this validation procedure to all measurements from all horses resulted in an overall sensitivity of 77% and an overall precision of 78% for the entire dataset. This performance is considered satisfactory for the quantification of mental state.

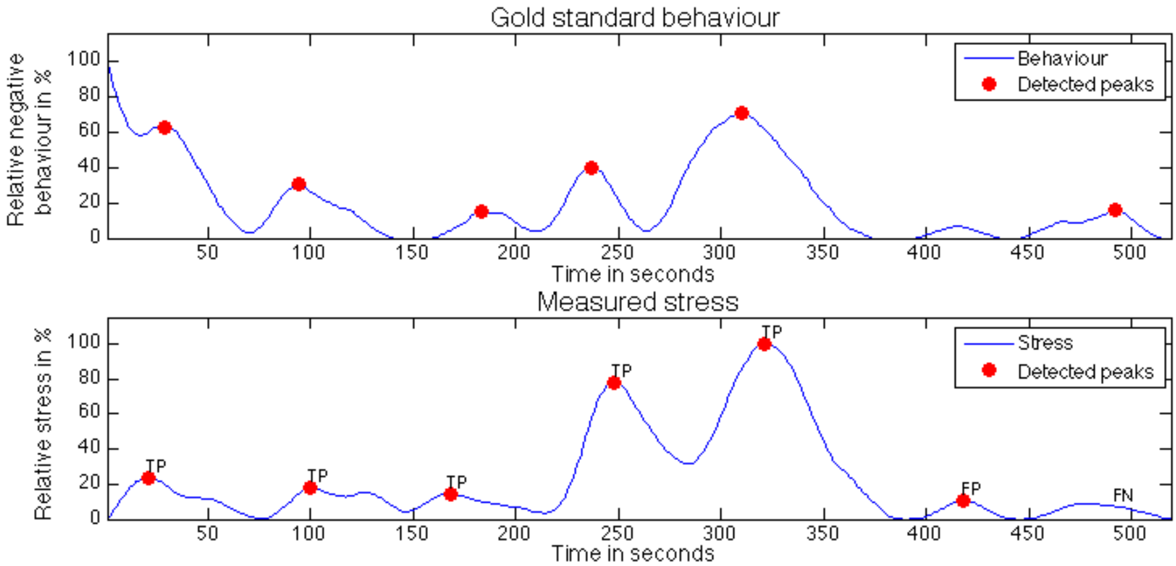


Figure 6: Peak detection and matching between the gold standard (top) and measured stress (bottom) for a good experienced horse during the OT. TP: True Positive, FP: False Positive, FN: False Negative.

In order to translate this mental state detection in an objective measure that can determine performance of police horses, two features were extracted from the relative stress graph of every horse for every measurement. The first one being the amount of times the horse's relative stress exceeds a certain threshold, further referred to as the #RT feature. This feature was determined for five relative stress thresholds RT1, RT2, RT3, RT4 and RT5, corresponding to 10, 20, 30, 40 and 50% of relative stress respectively. The zones in which exceeding of the threshold occurs, are called stress-zones. The second feature is the percentage of the time during which these thresholds are exceeded for the five different thresholds. This feature will be referred to %RT feature from this point on. Both the #RT and the %RT feature were stored for every measurement of every horse. **Figure 7** provides a visualisation of the stress-zones for the different thresholds by means of coloured bars.

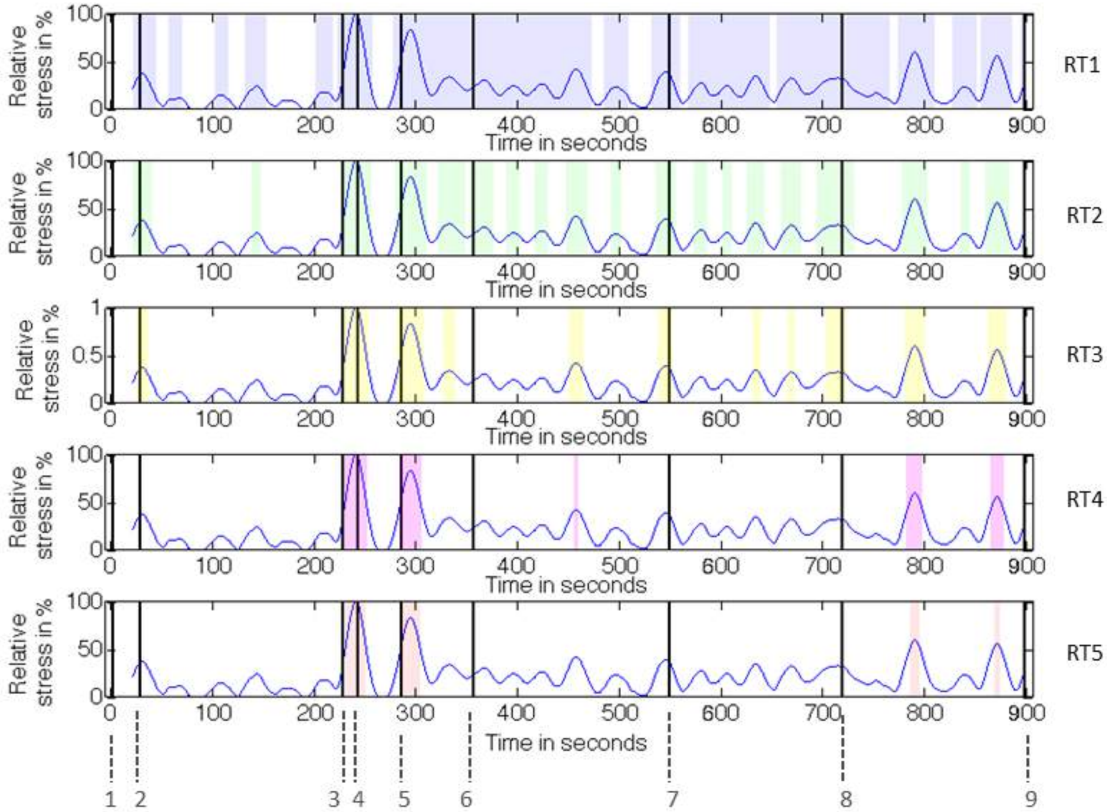


Figure 7: Indication of the stress-zones for the different relative thresholds for a good experienced horse during the CFW. The different steps in the CFW protocol indicated in the figure are: 1) Standing still, 2) Walking, 3) Rider dismounts, 4) Lead horse inside, 5) Attach horse, 6) Leave horse alone, 7) Take off saddle, 8) Leave horse alone, 9) End.

Results

In order to compare the features between the different horse categories, a multiple comparison test can be conducted. Since the sample sizes of the different horse categories are of unequal length, the Tukey-Kramer procedure was applied. In order to use this test, the data in the groups that are to be compared must be normally distributed and have equal variance (Rogan, Keselman, & Breen, 1977; Somerville, 1993). All feature values were tested for normality within each horse category using the Lilliefors test of normality (Scott & Stewart, 2001) with a significance level $\alpha = 0{,}05$. It was found that only the features #RT1, %RT1 and %RT2 were normally distributed. Testing whether these feature values have equal variance for the different horse categories was done using the Bartlett's test for equal variances with a significance level $\alpha = 0{,}05$ (Ma, Lin, & Zhao, 2015). It was concluded that all three features had equal variance. As a result, the Tukey-Kramer multiple comparison test was carried out for these three features using a significance level $\alpha = 0{,}05$. None of the horse categories were significantly different from each other for any of the features ($p>0{,}05$). Although no significant differences could be found, it could be observed that in general good beginner horses spend less time in the stress-zone defined by the first and second relative stress threshold compared to bad beginner horses. This is clearly visible in **Figure 8**. A possible interpretation is that good beginner horses spend less time in the distress zone than bad beginner horses.

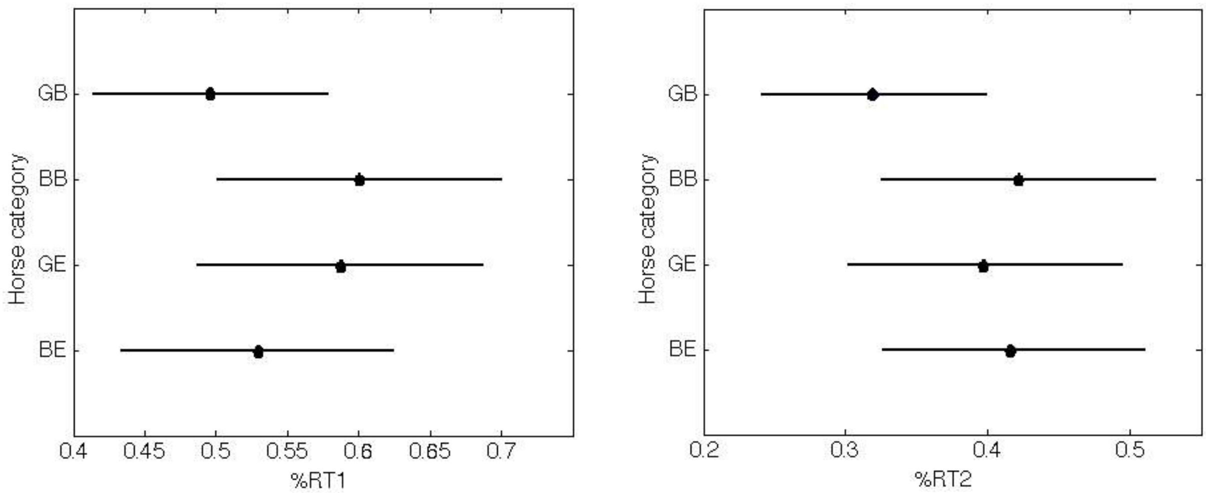


Figure 8: 95% confidence intervals for %RT1 (left) and %RT2 (right) for the four horse categories Good Beginners (GB), Bad Beginners (BB), Good Experienced horses (GE) and Bad Experienced horses (BE).

Comparison of the features between the different horse categories was also performed separately per protocol to investigate whether the feature values during a particular protocol might give some information on the suitability of a horse. Findings were that for the CFW protocol, a significant difference was found between good beginner and bad beginner horses for feature %RT2 ($p>0,05$). More specifically the values of the %RT2 feature are significantly lower for good compared to bad beginner horses. This means that during the CFW good beginner horses spend less time in the stress zone defined by RT2 compared to bad horses. This is in accordance with the previous observations about the comparison of the feature for all the protocols combined.

Conclusion

In conclusion, the goal of finding an objective measure that can be used to evaluate the suitability of horses as mounted police horse has been attained using mental state detection. This mental state detection can be fully automated and applied in real time, which can make it a useful tool for the mounted police to monitor the suitability of candidate police horses. However, the statistical results of this study should be interpreted with care due to the small sample size. As a final remark it must be noted that the concept of automated mental state detection in horses and their riders can be extrapolated to other equestrian disciplines. For instance it can be used to assess the interaction between the rider and the horse as a measure for rider-horse match. More importantly, automated stress detection in combination with stress management can help to improve performance and welfare of both the horse and the rider which might prove valuable both in recreational and competitive equitation.

Future work

Further validation on additional subjects is needed to increase statistical power. A future improvement for the automated mental state detection might be to distinguish between eustress and distress in the relative stress graph of the horses. In order to do so, a quantifiable and objective measure for performance will have to be defined to be able to link stress levels to performance levels. A second improvement could be to compute the relative stress of a horse during the measurement based on the horse's overall maximal stress level instead of computing it based on the maximal stress level during that particular measurement. In this way, it becomes possible to evaluate which stressor causes more stress in the horse. Based on this information it might be possible to predict the future performance of candidate police horses more accurately.

References

1. Westman, M., Eden, D. (1996), The inverted-U relationship between stress and performance: A field study, *Work and Stress* **10**(2): 165-173.
2. Peeters, M., Closson, C., Beckers, J. F., Vandenheede, M. (2013), Rider and Horse Salivary Cortisol Levels During Competition and Impact on Performance, *Journal of Equine Veterinary Science* **33**(3): 155-160.
3. Piette, D. (2015), Real-time monitoring of mental state in horses and their riders, Master's Thesis, KU Leuven, 116p.
4. MATLAB, R2014a(8.3.0.532), The MathWorks, Inc., Natick, Massachusetts, United States.
5. GH (2014). Figure of the human heart, <http://galleryhip.com/human-heart.html>. Accessed November 2014.
6. Jansen, F., Van der Krogt, J., Van Loon, K., Avezzu, V., Guarino, M., Quanten, S.,Berckmans, D. (2009), Online detection of an emotional response of a horse during physical activity, *Veterinary Journal* **181**(1): 38-42.
7. Young, P. C. (1989), Recursive estimation, forecasting and adaptive control, *Control and Dynamic Systems* **30**: 119-166.
8. Young, P. C. (2001), Recursive Estimation and Time-Series Analysis, New York, Springer, 504p.
9. Rogan, J. C., Keselman, H.J., Breen, L.J. (1977), Assumption violations and rates of type-I error for Tukey multiple comparison test – Review and empirical-investigation via a coefficient of variance variation, *Journal of Experimental Education* **46**(1): 20-25.
10. Somerville, P.N. (1993), On the conservatism of the Tukey-Kramer multiple comparison procedure, *Statistics & Probability Lettres* **16**(5): 343-345.
11. Scott, W.F., Stewart, B. (2001), Tables for the Lilliefors and Modified Cramer-von Mises Tests of Normality, *Communications in Statistics-Theory and Methods* **40**(4): 726-730.
12. Ma, X.B, Lin, F.C., Zhao, Y. (2015), An adjustment to the Bartlett’s test for small sample size, *Communications in Statistics-Simulation and Computation* **44**(1): 257-268.

Objective Measures of Eating Behaviour in a Swedish High-School

B. Langlet[1],A. Anvret[2],V. Papapanagiotou[3], C. Diou[3], I Moulos[4], C. Maramis[4], E. Lekka[4], R. Heimeier[5], A. Delopoulos[3] and I. Ioakimidis[1]

[1]Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden.
ioannis.ioakimidis@ki.se
[2]AB Mando, Stockholm, Sweden
[3]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
[4]Lab of Medical Informatics, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece
[5]Internationella Engelska Gymnasiet Södermalm, Stockholm, Sweden

Background

The importance of studying eating behaviour is recognized both in eating disorders (ED) and obesity (OB) [1,2]. However, the use of subjective, self-reporting measures in this fields is prevalent, even if they are known to be inaccurate across the general population [3] and even more so in people suffering from ED and OB (over- and under- reporting of their intake, respectively) [4]. On the other hand, studies which use objective measurements, such as calorimetry chambers and food double weighing, are far better at quantifying eating, but not in real-life conditions [5]. Additionally, most studies just report cumulative meal sizes, and/or estimated energy intake (kcal) [6,7], but fail to collect data on the specific patterns of eating, even though some eating styles (e.g., linear eating, see below) are more prevalent in people with ED and OB, and have been implicated in the onset of those conditions [8].

Retrospective age-of-onset data suggest that EDs usually occur in early adolescence [9]. Similarly, there is evidence that the distinct OB eating style develops early in life [10, 11]. Thus, any prevention intervention targeting the early normalization of the eating style of the individuals at risk for these two conditions should be implemented at an early age. This necessitates the development of measuring techniques which do not interfere with the users' behaviour in real life and at the same time allow the collection of objective data, always keeping in mind the age of the target population. Nowadays, wearable technologies and gadget-like sensor devices see increased mainstream use, become cheaper and more reliable, and they are inherently more acceptable among younger populations. Therefore, we believe, that the solution to objective behavioural recordings of eating and physical activity in everyday environments is the use of such innovative, user-friendly and appropriately validated devices.

With this goal in mind, we have envisioned and we are currently developing the SPLENDID; *Personalised Guide for Eating and Activity Behaviour for the Prevention of Obesity and Eating Disorders* system (through the homonymous EU project [12]), with the intention to create a personalised guidance system to help monitor and train children and young adults to improve their eating and physical activity by detecting subjects at risk for developing obesity or eating disorders and offering them behavioural guidance (through the use of non-invasive methods) in order to prevent the progression of these conditions [13]. In summary, the SPLENDID is a platform for the hands-off, objective behavioural monitoring, using the smartphone as the main input and info-collection hub, complemented by an array of external sensors; i.e., accelerometers for physical activity quantification, the Mandometer® [14] for eating behaviour monitoring during main meals and a newly developed chewing sensor, capable of automatically detecting snacking episodes in real-life settings [15]. All the sensor data is transmitted to the smartphone, where they are pre-processed and relevant behavioural indicators are then passed on to a central data collection and analysis server. The platform also includes web-based data access for the health professional or the researcher [13], and in the future will provide meaningful, personalised feedback to the end-user through the mobile. Note, that the specific SPLENDID use in the school includes real-life measurements of eating behaviour in the student population, used to identify adolescents with possible disordered eating patterns, acting as a first population screening for the full use of the system [13].

Traditionally, in a laboratory setting, the detailed quantification of the progressive changes of a subject's eating behaviour across a meal is measured through food weight-loss from the subjects' plate during the meal, by having a subject eat on a small weighing scale. Afterwards, the weight-loss data is transformed into intake-over-time and the new data series is fit with a quadratic equation (i.e., the cumulative intake curve): $y=kx^2+lx$ [16]. This is an elegant way to objectively quantify and dissect eating behaviour during a meal, in humans, and has led to the identification of two opposite eating styles in normal individuals, based on the change of their eating rate (decelerated and linear: with decreasing or stable eating rate, respectively), as the meal progresses [17, 18]. Studies in clinical populations have revealed that ED and OB individuals predominantly display a linear eating style, driving the hypothesis that a long-term linear eating leads to the disturbed eating styles associated with ED and OB [8]. Additionally, an individual's eating style is relatively stable, under similar conditions, across repeated meals [19], but it can be progressively modified by training [20], a principle that has been used in clinical practice for the treatment of ED [21] and OB [22], with promising results.

Until recently, this method involved a time-consuming process of data correction, including the manual subtraction of recording artefacts (e.g., cutlery placed on or removed from the plate) and the manual annotation of meal-related events (e.g., food additions). Currently, in clinical settings, the corrections are still made manually, whilst in research settings, a semi-automatic method of synchronising video and Mandometer® data has been developed [23]. Using such methods on a larger scale, i.e., in a school setting, we designed and implemented a non-invasive, reliable process, with the intention to: i) be acceptable by the majority of the students, ii) allow parallel data collection from a significant number of participants, iii) function within the time constraints set by the school (a lunch break in Sweden is usually around 30 minutes) and iv) assure minimum loss of data.

Study

Here, we present the first out of the four SPLENDID studies, planned to measure the detailed eating behaviour in student populations during school lunches, where we applied the methodologies traditionally used in a laboratory settings [23], in order to evaluate their use in a realistic school setting. In this case, we evaluate the feasibility of our methodologies based on: a) *the subjective ratings of acceptance from the participating students* and b) *an explorative comparison between the eating behaviour of male and female students*. We focused our explorative outcome analysis on sex differences (evaluated using t-tests), since this dataset is going to be used by the SPLENDID project, as a first reference base for eating behaviour characteristics in students during a school lunch. In short, the presented study aimed at proving both the real-life relevance and the intrinsic scientific value of the implemented methodologies, but also to provide a first estimation about eating behaviour values in the target population.

Ethics

All the presented processes were designed in coordination with the hosting high school, with the intention to be easy to integrate into the normal school life and have been approved by the Stockholm Regional Ethical Review Board. Written consent was obtained from all the participating students and their legal guardians.

Material and methods

Subjects: We recruited 41 out of 53 students from two classes of the Natural Science program in a Swedish high-school, located in Southern Stockholm. The students were between 16 and 17 years old (i.e., concluding the first year of the Swedish high school program; Table 1). The presented anthropometric measures were collected two days prior to the trials during a regular morning school lab session.

Mandometer®: All the meals were recorded using Mandometer® v4 (Mikrodidakt, Lund, Sweden), a weighing scale linked to portable computer [14]. The device records food weight reduction on a plate, over the course of a meal, at a sampling rate of 1Hz.

QR-code Android app: A research Android application was created specifically for the study, streamlining the process of coupling a participant with a random Mandometer® device on a random eating position. Additionally, the app allowed the recording of photographs of both the food on each plate and the participant's face, allowing the researchers to follow the participants in the recorded videos during the analysis phase (see below).

Table 1. Group characteristics of the participating students (n=41).

Mean (SD)	**Males**	**Females**
n	19	22
Age (years)	16.5 (0.5)	16.8 (0.3)
BMI (kg/m^2)	20.8 (2.1)	21.7 (2.8)

Study protocol: Both classes began their school day around 8^{00} am and followed their regular schedule until lunch, around 11^{30} am (the regular lunch time in Sweden). The lunch session protocols were identical between the two classes, ensuring that the students' daily habits were not affected. The food and the salads were served in a buffet-style (as is the habit in the specific school). The available food was identical between classes, and was provided by the school caterer. The 3 available main dish servings were: 1) beef patty 2) fish and 3) vegan patty. For all the dishes the side servings were: i) boiled potatoes, ii) brown or white sauce of choice and iii) the daily selection of green salads. All the foods represent food types frequently served in the specific high school. The choice of accompanying drinks were water and milk. Note that all the food (together with the salad) was placed on the same plate, following the practices of the school. Data recordings took place in a previously prepared and semi-attached part of the school cafeteria (*study room*) with 30 available *eating positions*. The serving area was videotaped by four cameras, used for external data and behavioural validation. Each *eating position* in the room was fitted with a Mandometer (see below), marked with a unique QR-code.

Figure 1. Pictures from the trial in the Swedish highschool (the Mandometers® v4 are visible):
A. The study room ready to receive the students for their lunch. **B & C**. Sample pictures of the lunches for two of the students, collected by a research assistant using the newly developed Android app. **D**. One of the classes having lunch (screen captured from a video recording).

Before the meal: 1) the participants were fitted with a personal QR code. 2) They served themselves from the nearby buffet (as usual), after which they returned to the *study room* and sat down at a random *eating position*. 3) research assistants coupled the students' QR-code with the QR-code of Mandometer® in their eating position, using a custom made Android mobile app. Afterwards, using the same app, a picture of the participant and a picture of the food on the plate was taken. Then, the students were able to eat their lunch using Mandometer®.

Finally, note that the students were allowed to visit the buffet and to add food as many times as they liked (as usual in this school) and then return and continue their meal using Mandometer®. No other constraints were set by the study protocol, allowing the students to follow their usual daily schedule when eating.

Subjective meal-related measures (e.g., hunger, fullness and sensory food characteristics) were collected during the lunch session, using analogue visual scales (generating scores of 0-100). However, since they are not the main focus of this study, they are not presented in detail. After the lunch session, the participants went on with their regular school schedule until the end of the school day, when, at around 4^{00}pm, they were asked to answer study- and equipment-related (e.g., evaluation of the mobile app design, comfort of participation in the study, self-assessment of possible behavioural changes due to the used equipment etc.) questions, using a 9-point statement-agreement scales (i.e., Select a point from *1:disagree* to *9:agree* for the sentence: "*I felt comfortable participating in the SPLENDID trial"*).

Data analysis

The video and Mandometer® data were transferred to a PC and analysed in semi-automatic, parallel fashion (the detailed methodology has been described and validated in detail elsewhere [23]). The video-corrected Mandometer® data were also used for the development of algorithms for the fully automatic Mandometer® weight-loss data correction [24]. In this case, the semi-automatic correction process is regarded as the golden-rule and it is the method used for the generation of the presented data. The corrected food intake data were used to calculate: a) the cumulative meal characteristics, i.e., the total food intake, the meal duration and the number and weight of bites, and b) the cumulative intake curve characteristics of the recorded meals.

Algorithmic correction: The fully automatic corrections of the Mandometer®-derived, weight-loss data were done by algorithms that remove recording artefacts (e.g., cutlery left on the plate and pressure put on the plate when handling the food) and identify the start and end of the meal, as well as spoonfuls and additions of food. First, the algorithm applies an adaptive pre-processing step which removes small artefacts. Subsequently, a Probabilistic Context-Free Grammar is used to interpret the meal recording as a series of events (bites, food additions, artefact placements and removals). Based on the detected events, all weight fluctuations not related to eating activities are removed, yielding a series of bite events that compose the food intake curve. Further details about the algorithm, as well as experimental results can be found in [24].

Statistical analysis: All the statistical analyses were made using R 3.2.3 (The R Foundation for Statistical Computing, Vienna, Austria). All the presented values are mean (SD) and only p-values <0.05 are reported.

Results

Feasibility: 40 out of 41 available datasets were successfully collected and analysed. 1 dataset lost due to equipment mishandling (*data retention rate > 97%*). Only the subjects with fully available data are included in the group comparisons. The entire protocol was completed without any delays during the allocated 30-minute lunch break in the school, causing *no disturbance to the daily school programme*. The student themselves felt comfortable participating in the study, with 100% of the students (n=41) rating their comfort more than 5 (in a 1-9 scale, with 9 fully agreeing to the statement *"I felt comfortable participating in the SPLENDID trial"*) and 73% of the students rated their comfort either 8 or 9 out of 9. No occurrences of disapproval/adverse behaviour were observed/reported to either the SPLENDID investigators or the school personnel. Similarly, the students thought that the study equipment did not affect their everyday behaviour, with more than 90% rating this effect less than 4/9. On the question of how willing are they to use the Mandometer® in a school setting, 31% agreed with the proposition on a rate higher than 5/9.

Sex differences in eating behaviour during the school lunch:
Cumulative meal characteristics: During this school lunch, male students *ate significantly more* than female ones; **456.5**±197.1 vs **307.18**±114.9 grams (**p<0.01**). However, the *average meal duration* in the two sexes *did not differ significantly*; **10.9**±4.8 vs **9.3**±3.7 min for males vs females. These data point towards a higher cumulative eating rate for the male compared to female students. Intake compared with control male vs. female; 525.5±182.7 vs. 298.8±104.0.

<u>Bite numbers and sizes:</u>Looking closer at bite characteristics allows explanation of the sex differences in eating behaviour. Thus, while male and female students took similar *number of bites* (**41.1**±18.6 vs **38.7**±17.9 for males and females), males took *significantly larger bites* (Figure 2) than females (**10.4**±1.8 vs **8.5** g, **p<0.03**). Bites compared with control 38.3±16.8

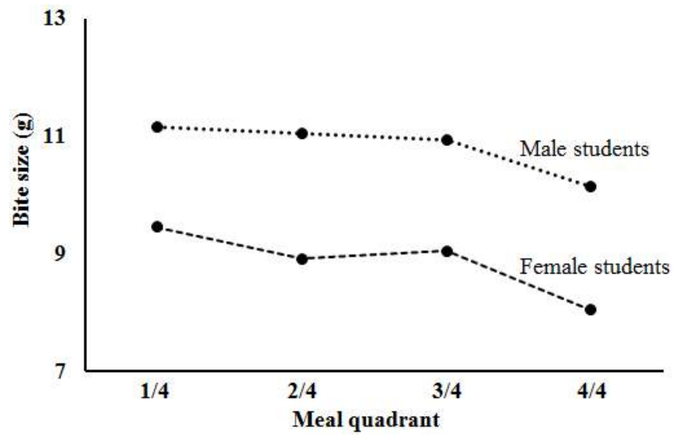


Figure 2. Mean bite size per meal quadrant (normalized over time),based on the recorded *male* vs *female* student data, during a lunch at a Swedish high school. 2-way ANOVA for repeated measures revealed that there is a significant effect of sex on the bite size (**p<0.05**).

<u>Cumulative intake curve:</u>The characteristics of the fitted quadratic equation (Figure 3) on the intake-over-time data, reveal that the overall *deceleration of the eating rate* in both groups was not significantly different (**-2.14**±2 and **-1.31**±1.9 for males vs females), but that males had a higher *initial eating speed* (**61**±21.9 vs **45.5**±21.4, **p<0.03**). Compared with control deceleration, -2.3±2.5. Compared with control initial eating speed, 60.2±26.3.

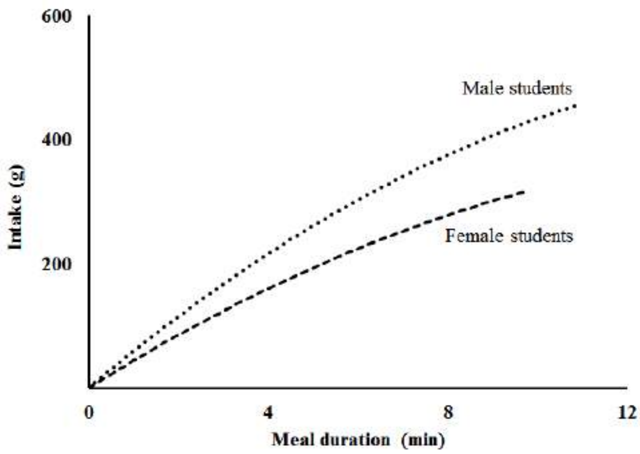


Figure 3. Schematic of the observed mean cumulative intake curves, plotted based on the recorded *male* vs *female* student data, during a lunch at a Swedish high school.

<u>Other lunch behaviours:</u>The male students *visited the buffet* more than the female students, but the difference was not significant (**0.6** vs **0.3** visits/person, ns). Finally, when the time of each individual away from the table (due to visit to the buffet, and/or getting more drinks, quantified through the video recordings), was redacted from the total meal duration, it was found that students spent on average **8.5 min** eating. This is similar to the meal duration of the control (average 8.5 min, SD 3.5).

<u>Subjective measures:</u> The two groups rated their *hunger* similarly *before the meals* (males vs females: **18.3** vs **28.1**) and rated their *liking* of the food similarly (males vs females: **56.3** vs **56.9**, ns), providing some evidence that the presented group comparisons are valid.

Combining the cumulative characteristics of the recorded meals, and the characteristics of the cumulative intake curves, it is clear that, *the male students ate more food than the females, during lunches of similar lengths (thus, showing higher overall eating rate across the meal). While both groups took similar number of bites, males took larger bites across the meal than females.*

Discussion

The presented results provide support to the notion that modern technologies can be used in real-life for the objective and detailed quantification of eating behaviour, even in challenging environments such as the school lunches analysed here. However, in order to apply methodologies used in semi-controlled settings (like a laboratory [23] or a clinical setting [8]), new supportive processes have to be developed and implemented, mostly in order to decrease the invasiveness of the recordings. In our case, the newly developed Android app allowed for quick, error-free and random assignment of individuals to random *eating positions*, not significantly upsetting the overall school habits of the participants. Additionally, the parallel development of automatic algorithmic processes for the data processing and analysis [24] will simplify similar trial methodologies even further. Once these processes are properly validated, they will allow recording of behavioural eating data of similar quality making cameras redundant and increasing their potential large scale use, thus allowing collection of representative, objective eating population data. In the future, such a system, complemented with some of the rapidly evolving algorithms for the automatic recognition of food components from the plate pictures [25], will be a very powerful tool for research on any type of eating behaviour measurements in real life.

The observed behaviours in this study are comparable with previously published data that were collected in laboratory settings. For example, in one study, 11 young women ate on average 264g [23], which is comparable to the 307g consumed by the younger female students in this study. Interestingly, the average number of bites taken across the meal was higher in the laboratory setting than in the school (51 vs 38 g) and the meals were longer (11.4 vs 9.3 min), even if they were eaten in absence of any social interactions. This difference is more pronounced once the observed "real eating time" per student is considered (8.5 min spent eating in average). In another similar study performed in a school setting, younger children ate similar quantities of food in even shorter periods of time [11]. This suggests that time constraints force young students into a "stressed" type of eating in the real-life school setting. In the school selected in this study, the students have a strict 25min time-window to get food, eat it and exit the dining areas, a typical time schedule for Swedish high schools. It is reasonable to hypothesize, therefore, that longer lunch times are needed for healthier eating behaviours. Interestingly, a similar conclusion has been recently reported, linking longer school lunch duration to healthier food choices [26]. This issue should be further researched, but it is becoming clearer that the perpetually shortening of school lunches is a cause of concern.

Although sex differences in eating behaviour in schools have been reported before [19], they have never been described in such detail. Interestingly, male and female students took a similar number of bites, males took larger bites than females, suggesting that bite size, rather than bite frequency, explains the sex difference in the cumulative food intake.

While the student self-rated comfortability and trial-process acceptance were already very satisfactory, the methodologies used here can be simplified and automated further, allowing detailed measurements in larger groups of people, in a wide variety of real-life circumstances. The data presented here provide a starting point for the future SPLENDID [12] studies aiming at developing a personalised guidance system to help children and young adults monitoring and improving their eating behaviour in challenging situations, such as the school lunch. That is especially important since school lunches often include time constraints forcing students to eat in ways that put them at risk of developing OB or ED [11].

Acknowledgements and statements

A big thanks to the administration, the teaching and the kitchen staff of the hosting school for facilitating every aspect of this study. Thanks also to the students for participating and to their legal guardians for allowing the children to take part. We really enjoyed working with you!

The data collection and analysis presented here has been supported by the SPLENDID programme (610746) of the European Union. B. Langlet, A. Anvret, and I. Ioakimidis have previously received/receive salaries by Mando Group AB, owner of the IPR for the Mandometer®.

References

1. DiLeone, R.J., Taylor, J.R., Picciotto, M.R. (2012). The drive to eat: comparisons and distinctions between mechanisms of food reward and drug addiction. *Nature Neuroscience* **15**, 1330-1335.
2. Walsh, B.T. (2011). The importance of eating behavior in eating disorders. *Physiology & Behavior* **104**, 525-529.
3. Körtzinger, I., Bierwag, A., Mast, M., Müller, M.J. (1997). Dietary underreporting: validity of dietary measurements of energy intake using a 7-day dietary record and a diet history in non-obese subjects. *Annals or Nutrition & Metabolism* **1**, 37-44.
4. Lichtman, S.W., Pisarska, K., Berman, E.R., Pestone, M., Dowling, H., Offerbacher, E., Weisel, H., Heshka, S., Matthews, D.E., Heymsfield, S.B. (1992). Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. *The New England Journal of Medicine* **27**, 1893-1898.
5. Seale, J.L., Rumpler, W.V. (1997). Comparison of energy expenditure measurements by diet records, energy intake balance, doubly labelled water and room calorimetry. *European Journal of Clinical Nutrition* **12**, 856-863.
6. Gianini, L., Liu, Y., Wang, Y., Attia, E., Walsh, B.T., Steinglass, J. (2015). Abnormal eating behaviour in video-recorded meals in anorexia nervosa. *Eating Behaviours*, 28-32.
7. Mooreville, M., Davey, A., Orloski, A., Hannah, E.L., Mathias, K.C., Birch, L.L., Kral, T.V., Zakeri, I.F., Fisher, J.O. (2015). Individual differences in susceptibility to large portion sizes among obese and normal-weight children. *Obesity (Silver Spring, Md.)* **4**, 808-814.
8. Ioakimidis, I., Zandian, M., Bergh, C., Södersten, P. (2009). A method for the control of eating rate: a potential intervention in eating disorders. *Behavior research methods* **41**, 755-760.
9. Nagl, M., Jacobi, C., Paul, M., Beesdo-Baum, K., Höfler, M., Lieb, R., Wittchen, H. U. (2016). Prevalence, incidence, and natural course of anorexia and bulimia nervosa among adolescents and young adults. *European child & adolescent psychiatry*, 1-16.
10. Faienza, M. F., Wang, D. Q., Frühbeck, G., Garruti, G., Portincasa, P. (2016). The dangerous link between childhood and adulthood predictors of obesity and metabolic syndrome. *Internal and emergency medicine*, 1-8.
11. Zandian, M., Ioakimidis, I., Bergström, J., Brodin, U., Bergh, C., Leon, M., Shield, J., Södersten, P. (2012). Children eat their lunch too quickly: an exploratory study of the effect on food intake. *BMC public health* **12**, 351.
12. SPLENDID. <http://splendid-program.eu>. Accessed 26 January 2016.
13. Maramis, C., Diou, C., Ioakimidis, I., Lekka, I., Dudnik, G., Mars, M., Maglaveras, N., Bergh, C., Delopoulos, A. (2014). Preventing obesity and eating disorders through behavioural modifications: The SPLENDID vision. *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on. IEEE*, 7-10.
14. Mandometer. <https://www.youtube.com/watch?v=d6KrYFqHnn0>. Accessed 26 January 2016.
15. Papapanagiotou, V., Diou, C., Lingchuan, Z., van den Boer, J., Mars, M., Delopoulos, A. (2015, September). Fractal Nature of Chewing Sounds. In New Trends in Image Analysis and Processing--ICIAP 2015 Workshops (pp. 401-408). Springer International Publishing.
16. Kissileff, H.R., Thornton, J., Becker, E. (1982). A quadratic equation adequately describes the cumulative intake curve in man. *Appetite* **3**, 255-272.

17. Meyer, J.E., Pudel, V. (1972). Experimental studies on food-intake in obese and normal weight subjects. *Journal of Psychosomatic Research* **16**, 305-308.
18. Westerterp-Plantenga, M.S. (2000). Eating behavior in humans, characterized by cumulative food intake curves—a review. *Neuroscience & Biobehavioral Reviews* **24**, 239-248.
19. Zandian, M., Ioakimidis, I., Bergh, C., Brodin, U., Södersten, P. (2009). Decelerated and linear eaters: effect of eating rate on food intake and satiety. *Physiology & Behavior* **96**, 270-275.
20. Zandian, M., Ioakimidis, I., Bergh, C., Södersten, P. (2009). Linear eaters turned decelerated: Reduction of a risk for disordered eating? *Physiology & Behavior* **96**, 518-521.
21. Bergh, C., Callmar, M., Danemar, S., Hölcke, M., Isberg, S., Leon, M., Södersten, P. (2013). Effective treatment of eating disorders: Results at multiple sites. *Behavioral Neuroscience* **127**, 878-889.
22. Ford, A.L., Bergh, C., Södersten, P., Sabin, M.A., Hollinghurst, S., Hunt, L.P., Shield, J. P. (2010). Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial. *BMJ* **340**, b5388.
23. Ioakimidis, I., Zandian, M., Eriksson-Marklund, L., Bergh, C., Grigoriadis, A., Södersten, P. (2011). Description of chewing and food intake over the course of a meal. *Physiology & Behavior* **104**, 761-769.
24. Papapanagiotou, V., Diou, C., Langlet, B., Ioakimidis, I., Delopoulos, A. (2015). A parametric probabilistic context-free grammar for food intake analysis based on continuous meal weight measurements. *Conference proceedings 2015, 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 7853-7856.
25. Kong, F., He, H., Raynor, H. A., Tan, J. (2015). DietCam: Multi-view regular shape food recognition with a camera phone. *Pervasive and Mobile Computing* **19**, 108-121.
26. Cohen, J.F., Jahn, J.L., Richardson, S., Cluggish, S.A., Parker, E., Rimm, E.B. (2016). Amount of Time to Eat Lunch Is Associated with Children's Selection and Consumption of School Meal Entrée, Fruits, Vegetables, and Milk. Journal of the Academy of Nutrition and Dietetics 116, 12.

Linked Data Sharing to Foster Consumer Based Science Enabled by Richfields: A Research Infrastructure on Consumer Health and Food

Karin L. Zimmermann[1], Muriel Verain[1], Marc-Jeroen Bogaardt[1], Anouk Geelen[2], Paul Finglas[3], Pieter van 't Veer[2], Monique Raats[4], Bent Egberg Mikkelsen[5] and Krijn Poppe[1]

[1]LEI Wageningen UR, the Netherlands. karin.zimmermann@wur.nl
[2]Division of Human Nutrition, Wageningen University, the Netherlands
[3]Institute of Food Research, United Kingdom
[4]University of Surrey, United Kingdom
[5]Aalborg University, Denmark

Abstract

There is growing interest in consumer health as related to food intake, other lifestyle behaviours and their determinants. However, data is fragmented, key information is lacking, and the resulting knowledge gap prohibits policy makers and companies to design effective public health nutrition strategies and reformulate food products. Making "the healthy choice the easy choice" requires knowledge on the context of personal lifestyle choices of citizens. To this end, technologies to measure and collect data (real-time and in situ), and methods and processes to align, link and share data on food-related behaviour and lifestyle are urgently needed to revolutionize research into the diversity of European consumers' behaviour, covering the whole range from purchase (e.g. in-store), preparation (e.g. in the kitchen) and consumption of foods (e.g. in the dining room). This will overcome the status quo where food-related consumer behaviour and lifestyle trends are studied in isolation, in short time frames and in a relatively limited social and physical context.

The aim of this paper is to give an overview of the Horizon 2020 project RICHFIELDS, that has started last year. RICHFIELDS aims at designing a RI on consumer behaviour and food intake, with a Consumer Data Platform, that is to a large extent based on consumer-driven data collection on food behavior and lifestyle. Consumers are central to the design: they harbour crucial information, as they increasingly adopt mobile apps and sensors, get access to e-business data and even medical information. Collectively, such real time and in situ data create new opportunities for research by e.g. monitoring food-related behaviour and providing personalized feedback. The RI Consumer Data Platform will enable scientifically sound inference by connecting consumer-generated data (open, big data, web based, unstructured) with established observational and experimental data.

Introduction

The EU H2020 Programme addresses healthy diets for the ageing European population, as well as the increasing relevance of environmental and social sustainability of these diets [1]. The Strategic Research Agenda and the implementation plan of the Joint Programming Initiative Healthy Diet Healthy Life [2] prioritized determinants of diet and physical activity to align research on healthy diet (and physical activity as part of lifestyle) choices in an encouraging societal environment. Regarding dietary behaviour, social sciences have emphasized the importance of personal characteristics that interact with the social environment at specific times and places. In this respect, a sound scientific evidence-base on consumer behaviour and lifestyle is crucial for research, governments and industries to adequately respond to the urgent health and sustainability challenges in the health and food domain.

Food-related consumer behaviour and lifestyle trends are studied in isolation, in short time frames and in a relatively limited social and physical context. Current data collection methods are expensive and show only certain aspects of behaviour (e.g. food purchase or intake) as they are typically based on interviews or recordings in surveys that sample sub-populations. These datasets are not sufficient to understand food intake;

not for companies to develop products for personal nutrition nor for governments for policy evaluation. Without more integrated data it will be difficult to make the healthy choice the easy choice.

To fill these gaps and advance food- and health related consumer research, which is relevant to the daily life of the majority of apparently healthy EU-citizens, new technologies to collect, align and share (real-time) data on food-related behaviour and lifestyle are urgently needed to revolutionize research into the diversity of European consumer behaviour, covering the whole range from purchase (e.g. in-store), preparation (e.g. in the kitchen) and consumption of foods (e.g. in the dining room). Consumer driven, big data and new tools to measure food intake are important drivers to bring consumer insights to a next level.

The FP7 EuroDISH project (Grant no311788, www.eurodish.eu,, [3]) has identified the need for research infrastructures (RIs) in the food and health domain that can advance research *within, among and over-arching* the so-called 'DISH' research domains: Determinants of dietary behaviour, Intake of foods and nutrients, Status and function of the body, and Health and disease prevention. The EuroDISH project's initial findings confirm the current disparate and fragmented health and food RI. Advanced RIs have been identified as useful to help facilitate health and food research (e.g., BBMRI-ERIC, CESSDA, EATRIS, ECRIN-ERIC, ELIXIR, EBI, ESS, MetaboHUB and SHARE-ERIC). However, none of these RIs can be considered health and food specific. Those "potential" or "developing" RIs identified that are health and food specific predominantly reside within one of the DISH model defined research areas "Intake", "Status" or "Health" as for the research area "Determinants" RIs are lacking [4]. Therefore the new H2020 project RICHFIELDS (Grant no 654280, [5]) aims to design a RI Consumer Data Platform with the focus on Food and Health Consumer Behaviour and Lifestyle. Based on all the data collected, the new RI will make it possible to create new knowledge in the 'DISH-domains' of Determinants and Intake, with close connections to the domains of Status and Health. In this way, the new RI will fill the present gap in the food and health domain and simultaneously connect consumer behaviour in the domains of Determinant and Intake to the largely disconnected research on Intake, Status and Health outcomes [3].The RICHFIELDS project is a 3 million euro project that will last for three years (2015 -2018) and has 16 European partners.

Aim

The aim of the RICHFIELDS project is to design a RI Consumer Data Platform which serves as an open access, distributed data-platform to collate and connect, and align and share innovative and existing data in order to enable researchers, policymakers and other stakeholders to develop, evaluate and implement effective food and health strategies both at the level of individuals and populations. The design of the RI Consumer Data Platform will function as a data platform that harbours a growing and flexible scientific evidence base for food and health behaviour and lifestyle of European Citizens considering state-of-the-art information and communication technologies for collecting big and open data created by consumers and researchers. It will integrate a large diversity of data that describes the interactions of time, place and individuals' behaviour, generated by machines, such as sensors gathering information, digital pictures and videos, purchase transaction records, GPS signals, etc. RICHFIELDS will design models to describe the provenance of data and explore mechanisms (like triple-level security, digital signatures etc.) to provide the security, the origin and the ownership of linked data.

Approach

The RI Consumer Data Platform will be achieved by integrating food and nutrient related data generated either for research purposes or for purposes other than research (e.g. commerce); either directly provided by consumers (e.g. uploading invoices, medical data) or indirectly by technology (e.g. apps, sensors) or information systems/processes (e.g. medical data, sales data, surveillance data). The combination of the dimensions of time (e.g. real-time, longitudinal data) and place (e.g., GPS, in situ data) will enable detailed description of the context and environments in which the behaviour is taking place, with a focus on three domains: purchase, preparation and consumption. The RI Consumer Data Platform will enable scientifically sound inference by connecting data and methods from consumer-generated data with established observational (epidemiology, surveillance, purchase) and experimental data (experimental kitchen, buffet, restaurant). The interaction with new technologies in combination with big data will bring food and health research and dietary surveillance to the next level. The RI Consumer Data Platform will be a network of data resources and services, distributed laboratories and experimental facilities, of participating legally autonomous organisations that will be working together. That sets all kind of requirements and preferences with regard to the ICT, business model, governance, legal and ethics issues.

The design of the RI Consumer Data Platform and its interfaces will address both functionality of food-related consumer behaviour and lifestyle as such as well as its viability and sustainability.

- At the data-level of the RI, the design will (1) provide the architecture based on 4 building blocks: data quality and assessment, meta data, standardization and linking) for a "big food related data" level, a platform of exchangeable consumer- and research-generated data that enables to extract and combine information on food-related consumer behaviour, and (2) develop and maintain standards for high quality data collection, either directly from consumers or indirectly by enrichment with existing available and unlocking new data sources.

- At the level of the viability and sustainability of the RI, the design will account for (1) flexible technical architecture to link individual and experimental data; (2) governance and ethical issues to support supply and use of data both in the development and maintenance phase; (3) a business model to underpin the feasibility of the RI (a service to stakeholders at government, SME & industry, research institutes, but also to consumers to manage their food and health data).

Gaps and needs

Within DISH (Determinants, Intake, Status and Health) areas, the research facilities, resources and services related to the *Determinants* area appear to be the most fragmented perhaps due to the less defined boundaries of discipline, paradigm, data source and methods of study [4]. Therefore, the RICHFIELDS project will be a step on the way to explore the sustainability and practicality of designing the RI Consumer Data Platform to specifically aid the study of the link between determinants and intake of diet-related lifestyle behaviours [6]. This includes exploring the collection of new data, re-using/collecting secondary data and investigating how determinants of diet-related lifestyle behaviour can be integrated into or extracted from existing health and food RIs. Research challenges and problems are listed in table 1as well as possible contributions of the RICHFIELDS project. Corresponding requirements and preferences towards the design of the RI Consumer Data Platform are formulated.

Table 1: Contributions of RICHFIELDS

Research challenges and needs	How RICHFIELDS project goes beyond the challenges and needs	Requirements & preferences for the design of the RI Consumer Data Platform
None of existing RIs can be considered food and health specific. There is a need for RIs in the food and health domain that advance research within, among and over-arching the DISH research domains. Current RIs on food-related consumer behaviour and lifestyle are adjudged to be in their infancy.	Greater co-ordination and support of food and health research. Using the enormous potential to create a RI to stimulate high-quality food and health research. Strengthen analytical capacity on food-related consumer behaviour and lifestyle.	The RI Consumer Data Platform serves as a data-platform, fills the present gap in the food and health domain, and connects consumer behaviour in the D-I domain to the largely disconnected research on I, S and H-outcomes.
The scientific community needs an advanced RI, relevant to the daily life of the majority of apparently healthy EU-citizens, industry and the scientific community needs.	An advanced RI that stimulates the development of unique services of added value to connect real- time in-situ data of the consumer to data of other RI's and experimental facilities.	Development of unique services for collecting, aligning and sharing innovative data types. Connections between consumer data with business and research generated consumer data.
Sound scientific evidence-based information on food-related consumer behaviour and lifestyle is crucial for research, governments and industries.	A RI that serves as a data-platform which performs the functions of a RI for researchers, industry and policymakers. The emerging knowledge will be accessible for public and private stakeholders.	Aspects as ICT solutions regarding data collection, big data storage, data standardisation (e.g., operating procedures, technical language and/or reporting of research/meta-data) and/or harmonisation (e.g., between MS) and data sharing, privacy, governance (e.g., management structure), sustainable funding, ethical/legal procedures), accessibility (e.g., to data, methodology, knowledge/expertise) and other relevant issues will be analysed and discussed.
Technologies to collect, align and share data on food-related behaviour and lifestyle are urgently needed. The current food and health research infrastructure is disparate and fragmented.	An architecture for a big data solution. The design will account for a flexible technical architecture.	The RI Consumer Data Platform will be designed by exploring concepts and tools for creating a semantic data model for linked (virtually integrated) structured, semi-structured, unstructured, big and open data, based on existing (standard) ontologies and incorporating aspects of research and business generated data as data provenance Cloud-based solution.
The status quo is that food–related consumer behaviour and lifestyle trends are studied in isolation, in short time frames and in a relatively limited social and physical context.	The RI will integrate a large diversity of data describing individual behaviour.	Integrating food and nutrient related data generated either directly by consumers e.g. apps or indirectly by technology (e.g., sensors) or information systems/processes (e.g. medical data, sales data, surveillance data).

Future perspective

Food and health are important issues for consumers (including patients). Developments in ICT (e.g., smart phones and cloud technology) help consumers to be more informed and manage their quality-of-life. These factors lead to more demand for science-based dietary advice. This trend will grow with more product reformulations (thanks to new options in plant production and food technology), more knowledge about the human body and brain (thanks to genetics, neuro-science etc.), and greater insights into the association between determinants and intake in relationship to food and health. The future perspective is to exploit this development by engaging with consumers and giving them an option to make data available to science, e.g. push the "blue button" or tic a "RICHFIELDS"-box in an app). Current data collection methods are expensive and show only certain aspects of behaviour (e.g. food purchases or intake) as they are typically based on interviews or recordings in surveys that sample sub-populations. These datasets are not sufficient to understand consumer behaviour; not for companies to develop products for personal nutrition nor for governments for policy evaluation. Without better data linkage it will be difficult to make the healthy choice the easy choice. Understanding of food-related consumer behaviour is crucial for preventive and health-promoting measures. The consumer-generated data (apps and sensors) provide the basis for development and evaluation of relevant behavioural feedback to consumers. Here, approaches jointly addressing the individual (health status, or from behaviourally or medically defined risk group) and his/her daily context (work, home, health, social, etc.) interact at the level of food purchase, preparation and consumption in the (super)market, the kitchen and during in/out of home consumption occasions. By giving consumers a platform through which they can manage their food and health data (of which they are increasingly the owner) RI Consumer Data Platform will make available data for researchers interested in studying food and health-related consumer behaviour. This data can be analysed using big data techniques [7] and be enriched through links with (open) datasets (The EU Open Data Portal *https://open-data.europa.eu/*). ICT technology can be applied that recognizes consumers as their own rich data source and create a possibility to manage it first of all for their own benefit but also to engage in (trans-disciplinary) science. Researchers using the RI Consumer Data Platform can as well use the data (subject to consumer consent), and can also ask the 'connected' consumers to test e.g., new ICT-apps for data collection. In addition, links will be created with experimental settings in kitchens, supermarkets and restaurants. This will focus particularly on the data and information management in such settings and the role of instant feedback from dietary and lifestyle advice on consumer behaviour.

Impact

Consumers share their individual data with the RI Consumer Data Platform on a voluntary basis although the EU Commission is currently reviewing its current legal framework on protection of personal data. The objectives behind this are to meet the challenges resulting from the use of new technologies and improve clarity and coherence of the EU rules for protection of personal data. Overlooking the privacy concerns, consumers' individual data create evidence-based and on-the-spot feedback loops on behaviour regarding purchasing, preparation and consumption. Also, data on the situation and consumption moment, location/GPS information, can be linked and becomes available on an individual level. The consumer will benefit from the platform by receiving individualized feedback on his provided data. Thus the RI Consumer Data Platform can contribute, in a scientifically and evidence based way, to trans-national and cross cultural consumer insights and feedback systems to represent food-related consumer behaviour and lifestyle more accurately in Europe.

Furthermore, the RI Consumer Data Platform will: (a) enable the research community to advance and innovate high quality and impactful research by means of a RI, and thereby, (b) enable the European research community to study food and health consumer behaviour and lifestyle in relation to purchase, preparation and consumption, both in real life, and in (semi-)controlled research settings and therefore support policy makers in private and public environments. In conclusion, the RICHFIELDS project aims to design a RI Consumer Data Platform for consumer behaviour and lifestyle, which performs the functions of RI for consumers, researchers, industry and policy makers in the domain of food and health.

Acknowledgements
The present paper is one of the results of the RICHFIELDS project Grant Agreement no 654280.

References

1. Horizon 2020 programme sections. Web-page link, last accessed Aug 2014: http://ec.europa.eu/programmes/horizon2020/en/area/food-healthy-diet
2. Joint Programming Initiative; Strategic Research Agenda 2012-2020 and beyond -Joint Programming Initiative, A healthy diet a healthy life (HDHL) JPI 2012. http://www.healthydietforhealthylife.eu/images/documents/JPI_HDHL_Implementation_Plan.pdf
3. Snoek, Harriette Dr, M, Rosalie Dhonukshe-Rutten, Ms, Kerry A Brown, Dr, Nadia Slimani, Dr, Paul Finglas, Dr, Giuditta Perozzi, Dr, Chris T Evelo, Dr, Ben van Ommen, Prof, Inge Tetens, Dr, Marga Ocké, Prof, Martine Laville, Dr, Leopold Fezeu, Dr, Adrienne E Cavelaars, Dr, Lada Timotijevic, Mrs, Karin Zimmermann, Mr, Krijn Poppe, Prof, Pieter van 't Veer (re-submitted 2015) EuroDISH: Assessing the Research, Societal, Innovation, and Policy Need for Research Infrastructures in the European Food and Health Domain, *Trends in Food Science & Technology*
4. Brown Kerry A, Lada Timotijević, Johanne L Arentoft, Julia Bardes, Astrid Böhm, Rosalie AM, Dhonukshe-Rutten, Léopold Fezeu, Paul Finglas, Marjolein Geurts, Camilla Hoppe, Jeppe Iversen, Martine Laville, Marga Ocké, Krijn Poppe, Nadia Slimani, Harriette M. Snoek, Inge Tetens, Pieter van 't Veer, Cécile Vors, Karin L. Zimmermann, (in revision 2016), Concepts and procedures for mapping food and health research infrastructure: New insights from the EuroDISH project, *Trends in Food Science and Technology*
5. Bogaardt, Marc-Jeroen, Anouk Geelen, Karin Zimmermann, Pieter van 't Veer, Monique Raats, Bent Egberg Mikkelsen, Krijn Poppe, Paul Finglas, Designing a Research Infrastructure on Consumer Health and Food Intake using E-science with Linked Data Sharing (RICHFIELDS), *International Journal of Behavioral Nutrition and Physical Activity* (to be submitted 2016)
6. Snoek, Harriette M, Lars MT Eijssen, Marjolein Geurts, Cecile Vors, Kerry A Brown, Marc-Jeroen Bogaardt, Rosalie AM Dhonukshe-Rutten, Chris T Evelo, Leopold Fezeu, Paul M Finglas, Martine Laville, Marga Ocké, Giuditta Perozzi, Krijn Poppe, Nadia Slimani, Inge Tetens, Lada Timotijevic, Karin Zimmermann, Pieter van 't Veer (submitted 2015, in revision 2016), Advancing food, nutrition and health research in Europe by connecting and building research infrastructures: Results of the EuroDISH project, *Trends in Food Science & Technology*
7. Mayer-Schönberger, V., & Cukier, K. (2014). Learning with Big data: The future of education. Houghton Mifflin Harcourt.

Measuring up to Expectations: A Food-related Study about ANS Responses and Expectations

Luz M. Verastegui Tena, Betina Piqueras Fiszman and Hans van Trijp

Marketing and Consumer Behaviour Group, Wageningen University, the Netherlands

Introduction

Understanding what drives consumers´ preferences and choices is of utmost importance for the food industry [1]. It seems that consumers' food choices are more unconscious than cognitive; which might explain why the tests commonly used by the food industry, such as liking tests and sensory analytical profiling, have a low predictive value with respect to product performance [2]. The responses of the autonomic nervous system (ANS) are believed to precede the conscious awareness of consumers and, consequently, are not susceptible to the bias often found in self-report measurements. These responses have been recently used to try to capture the consumers' affective reactions, such as liking and food preferences. To name some examples, He et al (2014) studied the physiological responses to pleasant and unpleasant food odours in a sample of 26 young participants and found higher heart rates when participants were presented to an unpleasant fish odour than a pleasant orange one [2]. Horio (2000) studied the heart rate changes when subjected to different tastes and found the highest increase with citric acid and a negative correlation between the increase in heart rate and the hedonic scales values in all solutions except for sucrose [3]. De Wijk et al (2014) studied the ANS responses to breakfast drinks and found a positive association between liking scores and heart rate [4].

Although evidence for autonomic specificity has been reported, the effects are inconsistent among studies. While many researchers agree on the fact that ANS responses are an important part of affect and emotion, there are still many non-affective and non-emotional states that also result in ANS activity, such as attention and mental effort. Moreover, emotions are adaptive responses to the world, not simply abstract sensations [5]. It comes as no surprise that it has not been possible to find ANS patterns as a reflection of emotional state. Due to this, researchers have suggested to study ANS responses in terms of broader dimensions, instead of in terms of emotions [6].

We believe that, in the food domain, expectation discrepancy might provide such a broad dimension. After all, sensory stimuli are always given a meaning by the consumers and evaluated according to their expectations [7]. A first necessary step, however, is to elucidate if ANS responses can be used as a measurement for expectation discrepancy or if they could be better used to represent another response linked to the stimuli. ANS responses are also involved in different processes in the body such as fight or flight responses, movement, breathing or the digestion of food [5]. The purpose of the present study is, therefore, to assess the validity of ANS responses for the measure of discrepancies in expectations.

Materials and methods

Eighty-nine healthy Dutch citizens ranging from 20 to 45 years of age were recruited from Ede area, the Netherlands, and surroundings. All participants read and signed an informed consent form before starting the test.

Heart rate and skin conductance were measured throughout the whole study with equipment developed by the Vrije Universiteit Amsterdam [8]. Seven sensor pads (five on the chest and two on the back) were placed on each subject. For the measurement of skin conductance two sensors were placed, one on the index finger and the other on the middle finger of the non-dominant hand. The researcher checked all signals before the start of the test.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

The test consisted of a tasting session and a presentation of the images of the main ingredients of the samples that were tasted. Participants were randomly allocated in one of two possible conditions. For both conditions, the explanation given was that they would try four different non-dairy drinks with a similar flavour but with different main ingredients. In reality three of the samples were the same (sweetened soy drink) and just one was different (rice drink). The rice drink became part of our design after pretests showed that adding a sample that was different, though similar in flavour, prevented participants from realizing they were tasting the same drink. Participants were instructed to taste the samples with their eyes closed, leave the sample in their mouth for 20 seconds and swallow it afterwards. To ensure that participants became familiar with drinking with their eyes closed they practiced first with a water sample. Before the start of the test, participants were asked to remain still, with their eyes closed and to breathe normally in order to have a baseline measurement of the ANS responses.

For Condition 1, participants were asked to first taste the sample and afterwards the researcher left the room while they were shown in a computer screen an image of the main ingredient of the sample. For Condition 2, the design was reversed; participants were first shown the picture before the actual tasting of the sample. All images were shown for five seconds and were preceded by a fixation cross that lasted 3 seconds. For the 3 samples that were the same, different images were shown as main ingredients: one neutral (soy bean), one positive/pleasant (chocolate), one negative/unpleasant (worms). The neutral image corresponded to the actual ingredient of the sample. In the case of the chocolate and worms image, we chose those images because they had been categorised as positive and negative stimuli in previous studies [8,9]. This would allow us to assess if the discrepancy in expectations had a different effect when the discrepancy was positive or negative.

Data treatment and analysis. Heart rate and skin conductance data were extracted with the VU DAMS programme version 3.9 [10]. All data were visually inspected for artefacts. For heart rate and skin conductance, seven-second labels were created which included the last second of the fixation cross, the five seconds the image was shown and the first second after the image was presented. As a first analysis, average heart rate (beats per minute) was measured for each label and skin conductance response amplitude (us) was calculated subtracting the minimum skin conductance to the maximum skin conductance from each label.

Image effect on heart rate and skin conductance (log transformed) was analysed by means of a mixed model Anova. Post hoc analyses were performed using Fisher´s LSD test. Analyses were done using the statistical software R version 3.2.2. For the analyses, only the results of the images that corresponded to the 3 samples that were the same were considered.

Results

Average heart rate was significantly different for each image ($p = 0.03$). There was no significant effect for type of condition ($p= 0.85$) or for the order of presentation of images ($p=0.40$). However, we found an interaction between the image and the order of presentation ($p=0.03$). In the case of skin conductance, the response amplitude was significantly different for each image ($p < 0.001$) but there was no significant effect for condition ($p = 0.73$) or order ($p = 0.47$). Similar to our results in heart rate, we found an interaction between the image and order of presentation ($p < 0.001$).

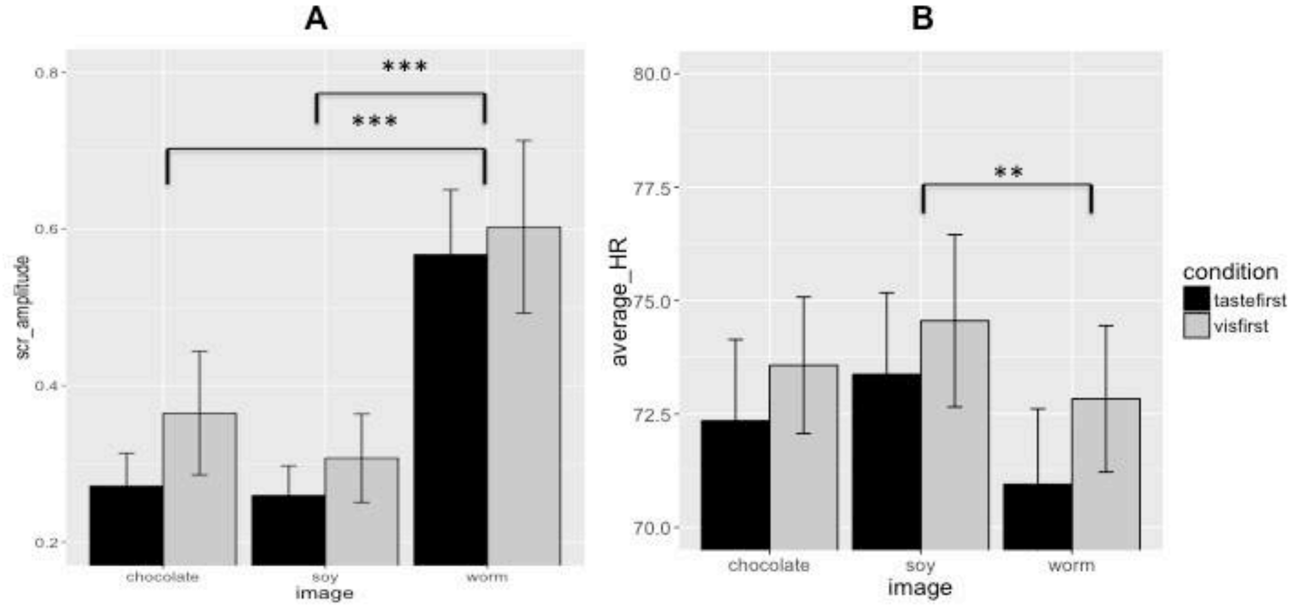


Figure 1 (A) SCR amplitude (us) and (B) heart rate average (bpm) during the image presentation before (gray) or after (black) tasting the sample.

Post hoc tests revealed that, for heart rate average, the main differences were between the worm and the soy image. For skin conductance, the main differences were between the worm and soy image and the worm and chocolate image (see Figure 1).

Discussion and conclusions

The main objective of this study was to validate the use of ANS responses for the measurement of expectation discrepancies. Even though we found a possible effect of expectation discrepancy, as shown by the significant interaction between the order of presentation and the images; we cannot conclude that discrepancy alone is responsible for this effect. It seems that other factors also play an important role.

For heart rate, we found a cardiac deceleration for the chocolate and the worms image, though the largest deceleration is for the worms. The experimental design and differences in stimuli account for the findings in cardiac response. Cardiac acceleration is usually found in tasks similar to social induction tasks, mainly because of the task´s demands and the intensity of emotions. When it comes to the perception of visual stimuli, however, the usual finding is a cardiac deceleration and the largest decelerations are found when viewing unpleasant stimuli that cause sadness or disgust[11] .

For skin conductance responses, we found that the worms image had a higher SCR amplitude than the soy image and the chocolate image. Skin conductance responses increase when viewing pleasant or unpleasant stimuli compared to neutral. They are only engaged with the most arousing pleasant or unpleasant stimuli which might explain why the SCR amplitude is higher for the worms image than for the chocolate image [12]. It is possible that participants did not find the chocolate image as arousing as the worms image.

As a conclusion, our results show that the changes in heart rate and skin conductance seem to be a response to the valence and arousal from negative or unpleasant stimuli, given the fact that the main changes in our study were between the worms and the other images. These findings can be used as guidance for the interpretation of ANS responses in consumer tests related to food products. It is necessary, however, to further study and disentangle the influence of valence and arousal in food-related ANS responses.

References

1. Clark, J.E. (1988). Taste and Flavour: their importance in food choice and acceptance. Proceedings of the Nutrition Society 57: 639-643
2. He, W., et al. (2014). Dynamics of autonomic nervous system responses and facial expressions to odors. Frontiers in Psychology. 5
3. Horio, T. (2000). Effects of various taste stimuli on heart rate in humans. Chemical Senses, 25: 149-153.
4. de Wijk, R.A., et al. (2014). ANS responses and facial expressions differentiate between the taste of commercial breakfast drinks. PLoS ONE, 9: e93823
5. Quigley, K., Lindquist, K. , and Feldman-Barret, L. (2014) Inducing and measuring emotion and affect. Tips,Tricks and Secrets, in Handbook of Research Methods in Social and Personality Psychology, H.T. Reis and C.M. Judd, editors. Cambridge University Press. p. 238-239.
6. Mauss, I.B. and Robinson, M.D. (2009). Measures of emotion: A review. Cognition & Emotion, 23: 209-237.
7. Lawless, H.T. and Heymann, H. (1999). Sensory evaluation of food: Principles and practices. Springer. p: 17-18
8. Genschow, O., Florack, A., Chib, V. S., Shimojo, S., Scarabis, M., & Wänke, M. (2013). Reaching for the (Product) Stars: Measuring Recognition and Approach Speed to Get Insights Into Consumer Choice. Basic and Applied Social Psychology, 35: 298-315.
9. Piqueras-Fiszman, B., Kraus, A. A., & Spence, C. (2014). "Yummy" versus "Yucky"! Explicit and implicit approach–avoidance motivations towards appealing and disgusting foods. Appetite, 78: 193-202.
10. de Geus, E.J.C., et al. (1995). Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. Biological Psychology 41: 205-227.
11. Palomba, D., Sarlo, M., Angrilli, A., Mini, A., & Stegagno, L. (2000). Cardiac responses associated with affective processing of unpleasant film stimuli. International Journal of Psychophysiology, 36: 45-57.
12. Bradley, M. M., & Lang, P. J. (2006). Motivation and emotion., in Handbook of Psychophysiology, J.T.Cacioppo, L. G. Tassinary, & G. Berntson, editors. New York: Cambridge University.p. 581-607

Adapting the eButton to the Abilities of Children for Diet Assessment

A. Beltran[1], H. Dadabhoy[1], T.A. Chen[1], C. Lin[1], W. Jia[2], J. Baranowski[1], G. Yan[3], M. Sun[2] and T. Baranowski[1]

[1] Department of Pediatrics, USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA. tbaranow@bcm.edu
[2] Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA, USA.
[3] Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA.

Introduction

Dietary assessment is fraught with error among adults [1] and especially among children [2, 3]. As much as 50% of child reported 24 hour dietary recall of foods for the previous day's breakfast and lunch had errors [4] as assessed against direct observation of consumption. Innovative technology may provide more accurate assessments of dietary intake [5].

One recently available innovative method is a camera worn on the chest (called an eButton) that takes images of whatever is in front of the wearer at brief intervals (2 sec to 10 sec) throughout the day [6-10]. Images of foods could minimize inaccurate recall of foods, but may have other limitations. These images can be clustered into images of a common single eating event [9-12]. Some of the foods in the food event image clusters can be automatically identified, but most must be identified by observers, usually dietitians [9, 10]. Food image portion size can be estimated using an innovative digital wire mesh procedure [9, 10, 13-16]. While the eButton has been tested among adults [14-16], adapting it for use by children requires understanding of its acceptance and procedural use. Two formative studies were conducted for using the eButton with children.

Study 1

Methods

Sample and Recruitment

In the Summer of 2014, a sample of 21 healthy 8-13 year old children and their parents were recruited from the Children's Nutrition Research Center's volunteer participant database, flyers posted throughout the Texas Medical Center in Houston, Texas, and online announcements on the Baylor College of Medicine and Children's Nutrition Research Center web sites. Of those 21 children 5 were siblings having a total of 16 parents. Saturation is a criterion for sample size in qualitative research. In our experience 16 would be ample to reach saturation. Inclusionary criteria were children between 8-13 years old (to extend our previous work [4, 17]), willing to eat one meal while wearing the eButton, and parent and child willing to answer questions about the eButton. In our previous research [4, 17], 8 to 13 years encompasses ages when children had difficulty reporting intake (8-9 years) to when no unusual difficulties were encountered (12-13 years). A detailed feasibility study protocol was approved by the Baylor College of Medicine Institutional Review Board. Signed informed child assent and parent consent were obtained. The parent received a stipend of $20 and the child received $15 for participating in this study.

Procedures

Two eButtons were tested: one light weight (about 42-45 grams) with an 8 hour battery life, and a heavier one (about 70 grams) with a 12-14 hour battery life. First, the heavier eButton was attached with a cloth strap on the front of the child's shirt at chest height and the child was asked to walk and run around the room. The same procedure was then followed with the lighter eButton. The child was then seated at a dining table. Height of the eButton camera to top of the table was measured. We asked if the child would wear the eButton for 2-3 days at a time and any concerns the child or parent might have wearing it. Lunch or dinner foods chosen by the child from

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

a menu were then placed on a standard dinner plate (diameter 26cm, depth 2.25 cm) or bowl (diameter 15.5 cm, depth 5cm) in front of the child. Eighteen menu specified items included, for example, a breakfast meal like oatmeal with fruit and yogurt; or a lunch selection like ham sandwich with a side of fruit and a drink. Once the child finished the meal, the child's thoughts about wearing the camera were queried.

A second staff person in a separate room interviewed the parent concerning the child wearing the eButton. Interviews were audio recorded.

Portion Size Estimation.

Portion size was estimated using wire-mesh software and images before and after intake from the pre-chosen meal. Technical details about portion measurement have been described [14-16]. True portion size before consumption was known from the detailed menu. Four analysts (one expert engineer from the University of Pittsburgh, who was facile in use of the wire mesh procedure, two research dietitians, and one undergraduate research assistant) estimated the foods portion sizes independently to assess interobserver reliability. The pictures were downloaded and imported to the specialized software (see Figure 1) [9]. Clear pictures were identified within the software before the foods had been eaten. The foods in the pictures were specified in the Food and Nutrition Database for Dietary Studies (FNDDS) food database embedded in the software. A 3-dimensional wire mesh, which best matched the shape of food, was selected, resized and deformed to fit the food (see Figure1). After fitting, the volume of the wire mesh (in cm^3) provided a measure of portion size. The diameter of the dinnerware (circular plate) on which the food was served was previously measured, providing a reference for the mesh. A total of 138 food items were analyzed. The portion sizes of 88 food items were estimated using the 3-dimensional wire mesh and 50 through visual estimation. Foods that were not able to be analyzed through the mesh software were either foods that were one whole piece with an irregular shape (e.g. individual pizza, hot dog) or the shape of a mesh was not available (e.g. meatballs in spaghetti, condiment packs). For these types of foods, a visual estimate of the portion size was entered directly in the software using the appropriate FNDDS options.

Data Analysis

Audio-recordings were transcribed; transcriptions were checked against audio-recordings; and imported into NVivo software (QSR NVivo 10.0, 2012, Doncaster, Victoria, Australia). Responses were thematically classified by question in the interview guide; codes were derived as the classification proceeded. Two coders coded all transcripts; differences in code assignment were adjudicated by consensus. Agreement on the portion size estimates among the four analysts and the known value were assessed using the intra-class correlation, which took into account observers nested within food items, nested within participants. The USDA's Food and Nutrient Database for Dietary Studies (5.0, 2012) was used to code dietary intake data and calculate calories and portion intakes [18].

Results

The average age of the 21 child participants was 10.4 (SD= 1.2) years. Other sample characteristics are in Table 1. Children reported neither consistent perceived difference nor preference between the lighter and heavier eButtons. The most common themes (20% or more of respondents) on the positive and negative aspects of wearing the eButton from children and parents are found in Table 2. The most common responses were the absence of perceived problems, but a few concerns were raised about comfort, wear during sports, attracting negative attention, and privacy.

	Study 1		Study 2	
	n=21		n= 12	
	Mean	SD	Mean	SD
Age (yrs)	10.48	1.24	10.67	1.5
	Frequency	Percent	Frequency	Percent
Gender				
Boy	16	76.2%	5	41.7%
Girl	5	23.8%	7	58.3%
Highest Education Completed by participating parent				
6th grade or less	1	4.8%	0	0%
High School graduate or GED	2	9.5%	1	8.3%
Technical school	2	9.5%	0	0%
Some college	7	33.3%	4	33.3%
College graduate	6	28.6%	2	16.7%
Post Graduate Study	3	14.3%	5	41.7%
Highest Education Completed in your household				
8th grade or less	1	4.8%	1	8.3%
High School graduate or GED	2	9.5%	0	0%
Technical school	4	19%	0	0%
Some college	5	23.8%	2	16.7%
College graduate	4	19%	4	33.3%
Post Graduate Study	5	23.8%	5	41.7%
Annual Household Income				
Less than $20,000	2	9.5%	0	0%
$20,000 to $39,000	3	14.3%	3	25%
$40,000 to $59,999	6	28.6%	3	25%
$60,000 to $79,999	3	14.3%	2	16.7%
$80,000 to $100,000	6	28.6%	1	8.3%
More than $100,000	1	4.8%	3	25%
Ethnicity				
Hispanic	10	47.6%	2	16.7%
African American	7	33.3%	9	75%
White	3	14.3%	1	8.3%
Asian-Non Vietnamese	1	4.8%	0	0%
Weight Status				
Underweight			1	8.3%
Healthy Weight	7	33.3%	6	50%
Overweight	6	28.6%	1	8.3%
Obese	8	38.1%	4	33.3%
Residence Type				
Single family house	16	76.2%	10	83.3%
Townhouse	1	4.8%	2	16.7%
Apartment	3	14.3%	0	0%
Other	1	4.8%	0	0%

Table 1. Demographics characteristics of children for Studies 1 and 2.

Child (n=21)	n	Parent (n=16)	n
Undesirable Aspects Identified		**Parents Concerns for using the eButton**	
Uncomfortable while wearing it, for example:	6	May fall or interfere while playing or during sports	10
- *Straight pin on back was uncomfortable*		Concern about attracting negative public's attention and kids asking about it	8
- *Magnet feels on body*		Privacy concerns while at home or school	8
- *Weight pulled the t-shirt down*		Concern about loss or damage	7
- *Strap felt rough, itchy, tight around neck*			
- *Strap didn't match school uniform*		**Desirable Aspects**	
eButton bounced when running or skipping	4	Parent not having problem charging eButton or uploading pictures for 2-3 days	15
Uncomfortable while wearing it during the meal:	9		
- *Worry it'll be damaged or get dirty while eating*		**Suggestions to improve eButton and overcome concerns**	
- *Restricted movement while eating or come in way*		Reduce size of eButton	7
Desirable Aspects		Have clear instructions of the process using the eButton	5
Wearing the eButton during a meal was comfortable and had no problems	14		
Fine to wear the eButton for 2-3 days at a time	15		
Suggestions to improve eButton			
Reduce size of eButton	6		

Table 2. Study 1 themes identified and responses to key questions.

Difficulty in obtaining pictures of foods was encountered with 7 of 21 children. For the three shortest children, a minimum height of 137 cm was needed to obtain images. Four children were too tall (≥156.7 cm) (camera to table measured 14 cm to 21.5 cm) since their images at their collarbone level were cut off. When the camera was lowered, the obtained images were acceptable.

Portion Size Estimation

Overall agreement on portion size estimation among the four analysts and the known value was moderate (ICC=0.60) when using 3-dimensional wire mesh and when estimating the portion size visually (ICC=0.57). The intra-class correlation was highest (ICC=0.97) for visual estimation when the engineer's values were excluded, likely due to the dietitians having increased knowledge of foods, volumes and portion sizes.

Changes in Method

Based on these results the following were changed for Study 2: the heavier longer duration battery eButton was selected; a soft paracord lanyard with emergency breakaway buckles and an adjustable double barrel sliding cord lock were added (see Figure 2) allowing movement up and down in case someone pulled on the string; the eButton was placed just below the collar bone (see Figure 3) to maximize obtaining clear images; a magnet was placed inside the child's t-shirt attached to the back side of the metallic eButton (to minimize bouncing); and the child was required to wear a t-shirt or high tank top to accommodate the magnet. No height restrictions were introduced since table heights at home would substantially vary. Given difficulties in identifying some foods and portions in the images (e.g. poor lighting, opaque packaging), a next day child food and portion verification procedure was added.

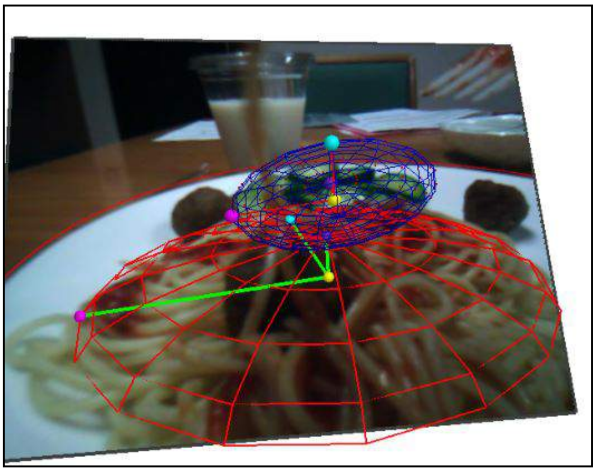


Figure 1. Half ellipse wire mesh for image volume estimation.

Figure 2. eButton sliding cord lock.

Figure 3. eButton location below the collar bone.

Study 2

Methods

Sample and Recruitment

Study 2 was conducted in winter of 2014-2015 having the child wear the eButton for one day at home and school and pilot testing the food and portion verification process. The same process was used to recruit 10 8-13 year old children (no siblings) and their parents. Ten was considered enough to encounter and identify the main problems in wearing an eButton all day. Inclusionary criteria were children between 8-13 years old, willing to wear the eButton for one day, and parent and child willing to answer questions about the eButton and send images via internet. Exclusionary criterion was participation in Study 1. A detailed study protocol was approved by Baylor College of Medicine Institutional Review Board, and written parent consent and child assent were obtained. Both staff were trained dietitians, who processed the Study 1 images, and passed all Baylor College of Medicine courses on human subjects and HIPAA confidentiality. Participating children were reimbursed $20 and parents $30.

Procedures

A full explanation was given to both child and parent in our Center on how to wear the eButton including type of shirt, specific location on the chest, turning it on/off, and charging the eButton. Staff provided a standard white plate (diameter 26cm, depth 2.25 cm), two size bowls (diameter 13.5cm and 15.5cm, depth 3.5 cm and 5cm respectively), and a small measuring tape; explained and demonstrated how to measure other dinnerware height and diameter (measures needed for volume estimation); demonstrated how to remove the Secure Digital card with the pictures to upload and send the pictures through bigfile; and how to complete a log of usage and problems. A food amount booklet (paper copy) was provided for the food and portion verification interview.

All images taken by the eButton were automatically encrypted, and unencrypted only by trained staff. Images with foods were identified and drawn into a separate file. All images with people in the file were de-identified by blurring faces or identifying information. Images not related to food events were deleted from the food analysis computer. After identifying food events, images were selected for before and after eating based on quality of the image, and whether the food could be clearly identified. When no clear image was available the best one was selected. Images with foods were uploaded to the volume estimation software. All the food items were searched in Food and Nutrition Database for Dietary Studies by dietitians within the volume estimation software and added to a list created by the software. When no exact food was found in the database a judgment of closest equivalent was made to identify the food. For example, "oatmeal not specified" was chosen versus "oatmeal with sugar, prepared with milk" since from the picture it was not possible to assess if it was prepared with milk or water or any additions to the oatmeal. Details on food preparation were asked at the food verification interview. Any unidentified food item, for example liquid in a colored cup or picture not clear, was flagged for food identification by the child. All the food and portion verification images were grouped in a file and sent via email to the parent before the verification phone call. On the same day the pictures were processed (the day after the images were obtained), a research dietitian conducted an audio-recorded phone-conducted food verification interview, where all the unidentified pictures were clarified, and the portion sizes were verified by the child. Food items recalled by the child as consumed, but not in the pictures, were also included._These data were collected and analyzed using Nutrition Data System for Research (NDSR) software version 2015, developed by the Nutrition Coordinating Center (NCC), University of Minnesota, Minneapolis, MN *[18].* After the food verification interview, child and parent interviews were conducted about their experience with the eButton.

Portion Size Estimation

Before the interview, portion size estimation using the 3-dimensional wire mesh software was conducted on the images able to be estimated. Foods that could not be software analyzed (e.g. plate size not measured) were visually estimated, and child-confirmed during the food and portion verification interview. After the interview, the list of food items was updated, and transferred back into the Food and Nutrition Database for Dietary Studies eButton software component.

Data Analysis

The data analysis procedures from Study 1 were used. The dietitians estimated portions independently.

Results

Twelve 8-12 yo children participated in this study. Average participant age was 10.7 (SD= 1.49) years. Details are in Table 1. Two children were dropped from the sample: for one the camera battery did not work properly and the family did not follow protocol procedures; the other did not have internet access at home and did not follow the wear time schedule. There was an average of 9 hours of images (minimum 4 hours; maximum 13 hours) from these 10 participants.

Food Identification and Portion Size Estimation

A total of 118 (84.2%) food items were identified before the food and portion verification interview (see Table 3). Twenty two (15.7%) additional food items were identified in the food and portion verification interview. Characteristics of these images are detailed in Table 3. After the food verification one food item was not recalled by the child even when the picture was available. This item was removed from the intra-class correlation analysis.

	Counts	%
Total food items	140	100
Total of food items before food verification	118	84.3
Total of food items added from food verification	22	15.7
Before meal pictures available (numbers of food items)		
Yes	114	81.4
No	26	18.6
After meal pictures available (numbers of food items)		
Yes	94	67.1
No	46	32.9
Identification of foods (numbers of food items)		
Foods accurately identified by dietitians	71	50.7
Blurry pictures	6	4.3
Dietitians not able to identify	33	23.6
More ingredient details needed for sandwich or mixed dish	14	10.0
More details needed for drink based on color	7	5.0
Not identified correctly (as determined by child interview)	9	6.4
Adequate placement of foods in the images (numbers of food items)		
Correct placement to obtain image	85	60.7
Not correct placement	47	33.6
Images not available	7	5.0
Identification of portion size (numbers of food items)		
Visual	124	88.6
Wire Mesh	16	11.4
Omission		
Picture is available but was not recalled by child	1	0.7
Meals recorded for the day (numbers of children)		
Complete meals (three meals + snacks)	5	50
Meals missing dinner pictures	4	40
Meals missing lunch & dinner pictures	1	10

Table 3. Food items before and after food verification in Study 2.

The intra-class correlation for portion size agreement between the two research dietitians using a consumed calories metric was moderate (0.53) with 3-dimensional wire mesh, but very good (0.98) when estimated visually. For the consumed servings metric, agreement was moderate when using the 3-dimensional wire mesh (0.59) and very good for visual estimation (0.99). When including the portion size verification data the agreement for consumed calories was both good using the 3-D wire mesh (0.62) and visual estimation (0.98). The approximate duration for the staff processing of the images for one day for one child was about 9 hours.

The most frequent child and parent themes about their experiences are summarized in Table 4. Again, the most common responses were in regard to ease and comfort of using the eButton.

Child (n=12)	n	Parent (n=12)	n
Undesirable Aspects		**Parents experience with eButton picture process**	
Attracted attention from other people (3 children felt uncomfortable about it and 3 didn't mind the attention, they liked it)	6	Instructions for uploading pictures easy to follow	9
		Average time to upload pictures = 30 min	8
Battery ran out or suddenly stopped	5	Parent thought the process for uploading was easy	8
Forgot to turn eButton back on	5	**Desirable Aspects**	
Difficult to wear it in school	4	Willing to do the upload process of pictures for 2-3days	7
Not secure enough	4	No concerns wearing for 2-3 days	5
Uncomfortable to wear	3	**Concerns for wearing the eButton**	
Desirable Aspects		Each concern mentioned by only one parent (unwanted attention, approval from school, size too big, child distraction, find appropriate clothes to wear eButton)	
Comfortable to wear during the day and meals	7		
No anticipated problems for wearing it 2-3 days	7		

Table 4. Study 2 responses identified to key questions

Discussion

While image assisted dietary recalls have been used for some time and have been accepted among adults and adolescents [19], this was a first attempt at using all day camera images to directly assess diet among children. These two pilot studies investigated whether 8-13 yo children would accept using a wearable camera, provide images for one day, and follow the procedures needed for direct dietary assessment. The eButton was able to be used by these children, but some children turned off the camera and forgot to turn it back on; the lighting made it challenging to see some of the images; some children ate standing up which did not allow a clear image; minimal necessary heights were identified to obtain images from a camera placed on a child's chest; and the battery lasted only 9 hours, so some of the meals were missed. Thus technical limitations persist in obtaining all day images.

Diet assessment with 24 hour dietary recalls among children has encountered 50% + error (intrusions and omissions) in food identification and related problems in portion size assessment [17, 20]. Smartphones and related software have been designed to eliminate or minimize this error by taking images of the foods before and after meals to permit the assessment of dietary intake from images [21, 22]. This method, however, requires that picture taking be volitionally initiated before and after the meal, the camera be at a certain angle with proper lighting, and a fiducial marker be properly displayed in the image, all of which may pose problems for children. While there was early hope that the image review process could be completely automated, thereby providing a time efficient and low cost method of diet assessment, humans have since been involved in food identification and portion assessment with images [22]. The promise of the eButton, and other all day image taking cameras [23], has been that the non-volitional nature of all day recording of images would identify all foods consumed; portion size estimation procedures could be applied to the images taken; and thereby lead to minimal error in diet assessment, possibly even done by non-dietitians. At this time, however, children's input is essential for food image verification to increase accuracy on food identification, especially on opaque cups for drinks and assembled foods, identification of missing foods (e.g. snacks), and portion estimation, especially for opaque packaged foods where volume can't be estimated from images.

Although intrusions and omissions in children's recall of foods eaten were no longer a problem when the images were visible, a number of technical problems existed. The size of the eButton needs to be smaller, thereby more discrete, so children can wear it comfortably. The resolution of the images needs to be improved in poor lighting since darker images are harder to identify. To make this system practical, the time to complete some of the image processing tasks needs to be reduced. Wire meshes for irregular, but common, food shapes (e.g. pizza slice, steak, hot dog) are needed. These problems are being studied by the engineers of our team who are designing new versions of eButton and software to improve performance.

The samples in both studies were small, self-selected, and unbalanced by gender, which are study limitations.

Conclusion

The eButton is a promising tool to minimize memory and portion size related error in child diet assessment, but poses new challenges. The eButton system for diet assessment may never be completely passive, requiring dietitian review of images and portions and child report of missing foods. Further development is needed to make the system practical and improve wearability and functionality among children.

Funding

This research was funded by a grant from the National Institutes of Health, National Cancer Institute (5 R21 CA172864), and institutional support from the U.S. Department of Agriculture, Agricultural Research Service (Cooperative Agreement 58-3092-5-001).

References

1. Thompson FE, Subar AF. (2013). Dietary assessment methodology. Edited by Coulston AM, Boushey CJ, Ferruzzi MG. *Nutrition in the Prevention and Treatment of Disease* 3rd ed. San Diego, CA: Academic Press: 5-46.
2. Livingstone MBE, Robson PJ, Wallace JMW. (2004). Issues in dietary intake assessment of children and adolescents. *British Journal of Nutrition* **92**, S213-222.
3. McPherson RS, Hoelscher DM, Alexander M, Scanlon KS, Serdula MK. (2010). Dietary assessment methods among school-aged children: Validity and reliability. *Preventive Medicine* **31**, S11-S33.
4. Baranowski T, Islam N, Baranowski J, Martin S, Beltran A, Dadabhoy H, Adame S, Watson KB, Thompson D, Cullen KW, Subar AF. (2012). Comparison of a web-based versus traditional dietary recall among children. *Journal of the Academy of Nutrition and Dietetics* **112**, 527-532.
5. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. (2010). Need for technological innovation in dietary assessment. *Journal of the American Dietetic Association* **110**, 48-51.
6. Sun M, Fernstrom JD, Jia W, Hackworth SA, Yao N, Li Y, Li C, Fernstrom MH, Sclabassi RJ. (2010). A wearable electronic system for objective dietary assessment. *Journal of the American Dietetic Association* **110**, 45-47.
7. Bai Y, Li C, Yue Y, Jia W, Li J, Mao ZH, Sun M. (2012). Designing a wearable computer for lifestyle evaluation. *Proceedings of the IEEE Annual Northeast Bioengeering Conference* **2012**, 93-94.
8. Li Y, Jia W, Yu T, Luan B, Mao ZH, Zhang H, Sun M. (2015). A low power, parallel wearable multi-sensor system for human activity evaluation. *Proceedings of the IEEE Annuals Northeast Bioengineering Conference* **2015**, 10.1109/NEBEC.2015.7117174.
9. Sun M, Burke LE, Mao ZH, Chen Y, Chen HC, Bai Y, Li Y, Li C, Jia W. (2014). eButton: A wearable computer for health monitoring and personal assistance. *Proceedings/Design Automation Conference* **2014**, 1-6.
10. Sun M, Burke LE, Baranowski T, Fernstrom JD, Zhang H, Chen HC, Bai Y, Li Y, Li C, Yue Y, Li Z, Nie J, Sclabassi RJ, Mao ZH, Jia W. (2015). An exploratory study on a chest-worn computer for evaluation of diet, physical activity and lifestyle. *Journal of Healthcare Engineering* **6**, 1-22.
11. Li Z, Wei Z, Jia W, Sun M. (2013). Daily life event segmentation for lifestyle evaluation based on multi-sensor data recorded by a wearable device. *Conference Proceedings IEEE Engineering in Medicine and Biology Society* **2013**, 2858-2861.
12. Li Z, Wei Z, Yue Y, Wang H, Jia W, Burke LE, Baranowski T, Sun M. (2015). An adaptive hidden Markov model for activity recognition based on a wearable multi-sensor device. *Journal of Medical Systems* **39**, 57.
13. Yue Y, Jia W, Sun M. (2012). Measurement of food volume based on single 2-D image without conventional camera calibration. *Conference Proceedings IEEE Engineering in Medicine and Biology*

Society **2012**, 2166-2169.
14. Jia W, Yue Y, Fernstrom JD, Yao N, Sclabassi RJ, Fernstrom MH, Sun M. (2012). Imaged based estimation of food volume using circular referents in dietary assessment. *Journal of Food Engineering* **109**, 76-86.
15. Chen HC, Jia W, Yue Y, Li Z, Sun YN, Fernstrom JD, Sun M. (2013). Model-based measurement of food portion size for image-based dietary assessment using 3D/2D registration. *Measurement Science and Technology* **24**, 105701.
16. Jia W, Chen HC, Yue Y, Li Z, Fernstrom J, Bai Y, Li C, Sun M. (2014). Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera. **17**, 1671-1681.
17. Diep CS, Hingle M, Chen TA, Dadabhoy HR, Beltran A, Baranowski J, Subar AF, Baranowski T. (2015). The Automated Self-Administered 24-Hour Dietary Recall for Children, 2012 version, for youth aged 9 to 11 years: A validation study. *Journal of the Academy of Nutrition and Dietetics* **115**, 1591-1598.
18. Feskanich D, Sielaff BH, Chong K, Buzzard IM. (1989). Computerized collection and analysis of dietary intake information. *Computer Methods and Programs in Biomedicine* **30**, 47-57.
19. Gemming L, Utter J, Ni Mhurchu C. (2015). Image-assisted dietary assessment: A systematic review of the evidence. *Journal of the Academy of Nutrition and Dietetics* **115**, 64-77.
20. Baranowski T, Islam N, Baranowski J, Cullen KW, Myres D, Marsh T, de Moor C. (2002). The Food Intake Recording Software System is valid among 4th grade children. *Journal of the American Dietetic Association* **102**, 380-385.
21. Schap TE, Zhu F, Delp EJ, Boushey CJ. (2014). Merging dietary assessment with the adolescent lifestyle. *Journal of Human Nutrition and Dietetics* **27 Suppl 1**, 82-88.
22. Martin CK, Nicklas T, Gunturk B, Correa JB, Allen HR, Champagne C. (2014). Measuring food intake with digital photography. *Journal of Human Nutrition and Dietetics* **27 Suppl 1**, 72-81.
23. Gemming L, Doherty A, Utter J, Shields E, Ni Mhurchu C. (2015). The use of a wearable camera to capture and categorise the environmental and social context of self-identified eating episodes. *Appetite* **92**, 118-125.

The Potential of Enriching Food Consumption Data by Use of Consumer Generated Data: a Case from RICHFIELDS

Jos van den Puttelaar[1], Muriel C. D. Verain[1] and Marleen C. Onwezen[1]

[1]**LEI Wageningen UR, the Netherlands. jos.vandenputtelaar@wur.nl**

Introduction

The relation between food and health is a topical issue, due to all kinds of food-related diseases that are omnipresent in today's society. Accurate insights in food behaviour are difficult to obtain, but are essential in understanding people's food consumption and the relation with health. Understanding food choice is therefore of significant interest to policy makers, nutrition scientists as well as food producers and retailers.

Recent developments in ICT provide new avenues to facilitate data collection on food behaviour. Surveys are more and more often web-based, which saves the researcher time and facilitates access to respondents. In addition, a wide range of apps have been developed to improve the measurement of food intake for example (e.g. apps for food diaries or short FFQs). Beyond time-savings and accessibility advantages, another important advantage of using apps is the possibility to collect the data real-time. Consumer-generated real-time data, through the use of mobile apps or tech-wear, provide a wealth of information on food choices in a real-life context. In the RICHFIELDS project – a three-year European funded project – innovative consumer support tools to collect purchase, preparation and consumption data real time and on the spot will be the focal point. This project will design a world-class infrastructure for innovative research on healthy food choices. Existing data on food products, food intake, lifestyle and health will be linked, including data coming from traditional sources as well as from the more innovative tools with consumer-generate data. The aim of the current paper is to present one such innovative tool, the WUR Food Intake App (working name), and to provide insight in how such a tool can overcome several drawbacks that the more traditional data collection methods in food and health research face.

Traditional research methods

The research into food consumption tends to focus on either dietary intake or socio-psychological determinants rather than a combination of both research domains. Combined insights connecting both domains is often lacking. In addition, the traditional data collection methods in both research domains face a range of drawbacks, that overlap to a large extend. First, data collection forms a large burden for the respondent. Second, measurement error is a problem. Third, the data collection process is time-consuming. Forth, there is a lack of insights into consumption patterns and changes thereof. And fifth, the source of information is often limited and can be much richer. Throughout the remainder of this paper we will discuss the five aforementioned drawbacks and how these can be overcome by the use of new technologies and innovative combinations of disciplines.

Drawbacks within fields of study

Dietary intake. 24-hour recall methods, food frequency questionnaires and diet histories are currently the most popular methods to investigate food intake [1] [2]. These methods have a number of drawbacks. The severity of the drawbacks depend on the chosen method but can be summarized as follows: 1) a large burden for the respondent, because food consumption data has to be entered in detail and for a longer period of time to get a complete overview of the diet; 2) measurement error of the intake due to the large time span between the moment of consumption and the moment of intake registration (the so-called recall bias) [3], because consumers often forget what they have eaten during a day; 3) a long time lag between the research question is formulated and the moment that the data is available, due to the time it takes to collect respondents, to collect the data, and to make the data workable; 4) a lack of insights into changes in consumption patterns and meal composition; 5) and a limited source of information, due to a narrow focus on food intake without including data on contextual factors (e.g. consumption moment, consumption location, or the social environment) or determinants of intake (e.g. food choice motives).

Socio-psychological determinants. Social scientists who collect data on determinants of food choice as well as on contextual factors are confronted with similar drawbacks. Socio-psychological determinants are traditionally collected with questionnaires or in experimental designs. The drawbacks of these methods to a large extend overlap with the drawbacks of the traditional methods for dietary intake. These drawbacks refer to: 1) a large burden for the respondent, because of lengthy surveys or time-consuming experiments; 2) a social responsibility bias, because consumers intend to 'please' the researcher in answering questions; 3) a long time span to collect the data, due to the time it takes to collect respondents and retrieve the data; 4) a lack of behavioral measures, real time and on the spot, which is necessary to relate food consumption determinants to actual behavior; 5) and a limited source of information. More specific to this last drawback, traditional lab- and natural *experiments* have provided a crucial understanding of food consumption, but a major issue of these studies is that they often only provide insights for a specific context, moment or product(group). On the contrary, validated *questionnaires* on socio-psychological variables, such as the food choice questionnaire (measuring food choice motives) [4], are often general and miss the link with specific situations, products or contexts. An overview of de drawbacks of the traditional methods is presented in Figure 1.

Drawbacks due to integrations between fields of study

Next to the drawbacks linked to the collection of intake data or socio-psychological data, a major issue in current food research is the scattered data and limited possibilities to link different types of data from different sources [5]. Food choice motives (e.g. importance of health considerations in people's food choices) that people report can for example not be related to actual consumption. This lack of data linkage forms a barrier to obtain a clear understanding of food behaviour [6].

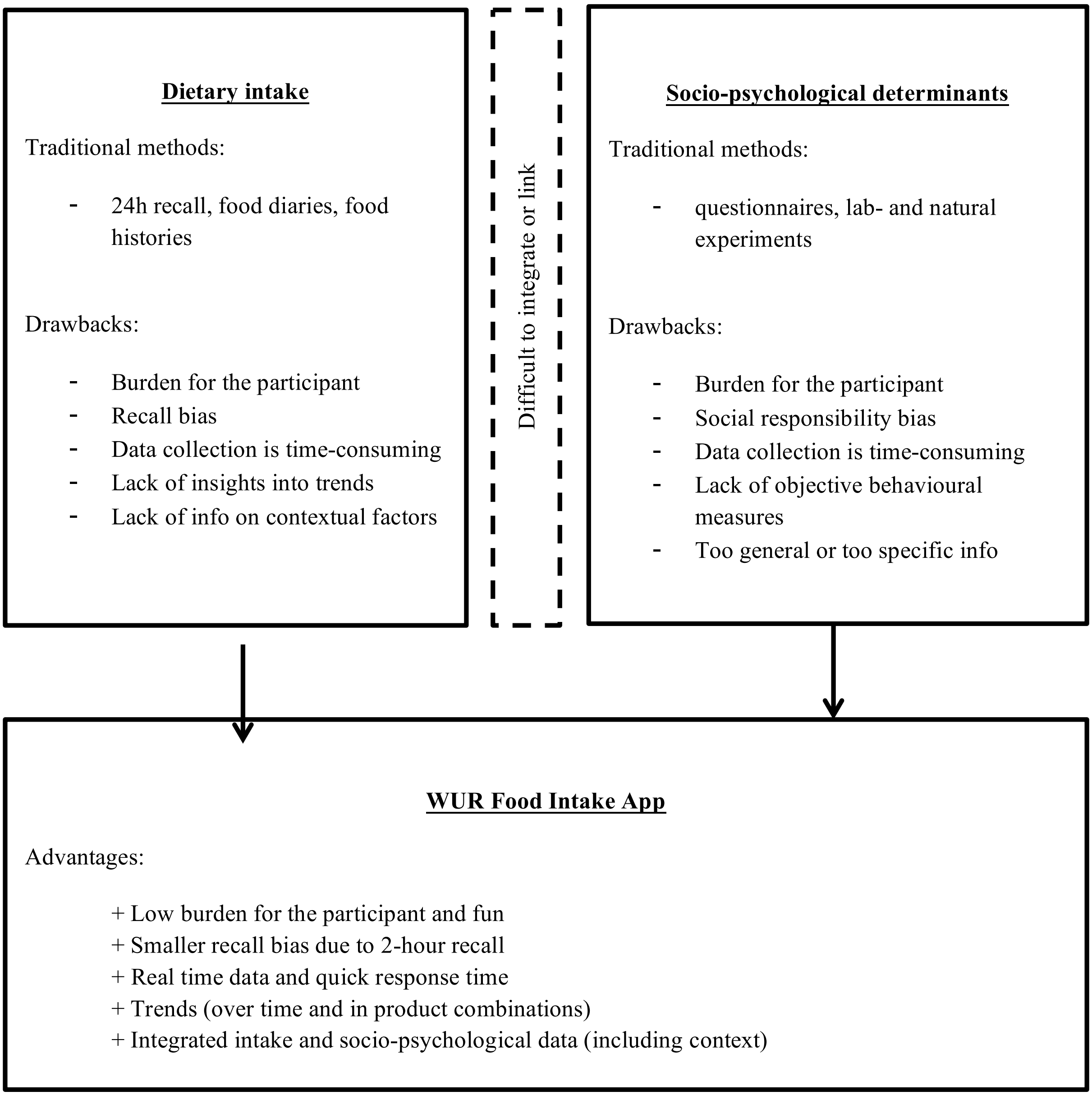


Figure 1. Generic model of food intake methods

The Food Intake app

The Food Intake App is developed with the purpose of complementing the traditional research methods outlined above, by overcoming the major drawbacks. The Food Intake app is a food consumption app that combines real world context, time series and socio-psychological insights (e.g. food choice motives). The app uses multiple measurement moments, resulting in a low cognitive load for the respondent. Compared to the traditional 24-hours recall method to measure food intake, the advantage of the app is the 2-hour recalls with random intervals. This lowers the burden for respondents (1) and reduces the recall bias (2). The recalls are not only spread in time, but also over participants. A user will thus not have to collect data continuously in a short time period, but rather fragmented over a longer period of time. By aggregating all the snapshots of food consumption, new insight can be achieved in trends in consumption patterns and product combinations on specific moments or in specific context. Due to the consumer-centred development, the use of the tool is not only low burden, but also fun and interesting (1). The app is designed in such a way that a user would like to keep participating without interfering with the data itself. There are several psychological techniques to keep people playing [7]. First of all, the interface has a more than functional design (e.g. attractive), users can build up a profile on socio-psychological data and food-intake, furthermore they receive feedback on their consumption pattern and interesting facts and figures about food. In this way, app-users should stay motivated to occasionally keep entering their food intake data, resulting in a large dataset of real-time data. The fact that consumers regularly and voluntarily fill out their consumption data makes the data collection quick. When a specific research question is formulated, users can receive a pop-up with the question to enter certain data, and they can do that real time and on the spot (3). This way of data collection also makes it possible to focus questions on certain countries or consumer groups, and to compare data over countries and consumer groups. Another advantage of consumers filling out data regularly over longer periods of time is the possibility to detect trends. Both trends in changing consumption patterns over time, and trends in product combinations on certain consumption moments can be investigated (4). The data resulting from the app is mainly beneficial for researchers, who can use the wealth of data on intake, motives, and contextual factors to improve the understanding of food behaviour (5).

The innovative nature of the food intake apps lies in the fact that a consumer voluntarily collects his or her own data, real time and on the spot. This method relates to new developments of quantifying user generated data, like shown in extremes by the Quantified Self movement (Quantifiedself.com), researcher self-measurement generated data like the Nutrition Researcher Cohort (https://humanstudies.tno.nl/nrc), motivation and health data (www.thepreciousproject.eu), and feedback on nutrition and health (http://www.qualify-fp7.eu). These studies have in common that they focus on specific (bio)markers or nutritional components. A consumer-focused real time food intake tool with a low burden will provide new insights.

The aim for a consumer-oriented intake app enriched with data of socio-psychological and contextual factors resulted in the design of the front-end (an appealable and easy app) and a back-end (connecting data that researchers can use). The app will be further tested and validated in a reference sample. In chronological order the following events are programmed: A participant receives at a randomized time a question if he or she ate something in the last 2 hours. If so, the next screen shows a dinner table. In this part the participant can choose the type of consumption moment (e.g. breakfast) and can enter information on what kind of food was consumed in the last 2 hours. The food products can be searched by typing in the search column or by clicking on the right category (e.g. bread) and then the right product (croissant) and the amount (e.g. 2 pieces). These few steps are the main part of the data collection. Further enrichment of the data is achieved by building up the profile. At any moment the participant can check the progress of his or her profile in relation to either demographics or food intake. A few examples are given in Figure 2 in which the choice for consumption moment, intake and lastly the progress of the profile is shown. Users are encouraged to reach the finish line as the progress is shown by a double helix showing more progression as you answer more questions. Furthermore, as the participant answers more questions, more facts and figures are given and finally feedback can be generated.

The back-end of the app consists of the products consumed, connected to an anonymized participant number. Next to the product consumed also the demographics of the participant, time of consumption (in a 2-hour interval) and the NEVO code (Dutch food nutrition database) of the product in question are automatically added. Especially this last part is important to be able to analyse the consumption pattern in terms of nutrition, and to be able to compare it to other databases. An added value of this way of data collection is to be able to get insights in product combinations that are regularly consumed together. Finally, such a design gives the freedom to add socio-psychological data in the form of questionnaires for example.

The data is stored on the server of LEI – which is an independent research institute. Private partners do not have access to personal data and data is only made available on an aggregate level. Before participation, the participants are informed about the data policy and asked for their consent.

Figure 2. Screenshots of the WUR Food Intake app.

The front-end of the app was designed by the software company Ranj (www.ranj.com) together with LEI-Wageningen UR and consists of the following parts (see Figure 2):

Validation and development

At this moment, developments are still in progress. The development is considered a continuous loop of feedback and improvement relating to both consumer usability of the app and usability for research. In terms of the technology readiness level, we are now a system prototype demonstration in the operational environment phase (TRL 7) [8]. At the same time, studies to validate the measurement technique are being developed. Validation will be done by a selected panel of consumers. They will receive a version of the app that replicates the 2-hour recall method by splitting a day up in 12 parts. Respondents receive reminders until 3 full days are filled. Their intake data will be compared with traditional methods like the food frequency questionnaire and other 24 hour recalls to validate the novel 2-hours recall research method. It is of importance to keep in mind, though, that the aim of the app is not to measure food intake more accurately in terms of specific types of products and portion sizes, as compared to traditional methods. The aim is to get clear insights in dietary patterns: which types of products are consumed at what moments in time, and what product combinations are usually eaten. Further development of the app by including more accurate intake measurements could be a next step, but is not the focal point at this first stage. In the future, the app could be extended by included features that measure intake in a different way, such as photo capturing features or barcode scanning. These features are not sufficiently developed yet. Photo recognition faces inaccuracy with portion sizes estimations and product recognition, and barcode scanning is often a problem due to incomplete databases. In addition, some types of products (e.g. fruits and vegetables or composed meals) do not always have barcodes. Therefore, registration of consumption by omitting manual registration of product intake is not an option yet, and therefore inclusion of pictures and barcode scanning will be an additional burden for the consumer. Future extensions of the app in this direction, when these types of features are further developed, could possibly make consumption registration more accurate, easier and faster on the long run.

Conclusion

The reason to start the development of an intake tool was to collect food intake data in an easy way for the consumer as well as for the researcher and to enrich intake data with socio-psychological data and insights in food consumption trends. The resulting WUR Food Intake App overcomes a range of drawbacks of traditional data collection methods in food intake and socio-psychological determinants. In addition, the app integrates data on determinants and on actual consumption. Ultimately, the data that is collected by the app can provide insights leading to a better understanding of food consumption. This improved understanding is useful in changing consumption behaviour towards more healthy food patterns.

Reference

1. Ocke, M.C., Bueno-de-Mesquita, H.B., Goddijn, H.E., Jansen, A., Pols, M.A., van Staveren, W.A., & Kromhout, D. (1997). The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. *International journal of epidemiology*, **26**: S37.
2. Thompson, F.E., & Subar, A.F. (2008). Dietary assessment methodology. *Nutrition in the Prevention and Treatment of Disease*, **2**: 3-39.
3. Subar A.F., Freedman L.S., Tooze J.A., et al. Addressing current criticism regarding the value of self-report dietary data. *The Journal of Nutrition 2015*, **145**: 2639-45.
4. Steptoe, A., Pollard, T.M., & Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: The Food Choice Questionnaire. *Appetite*, **25**: 267-284.
5. Snoek, H.M., Dhonukshe-Rutten, R.A.M., Brown, K.A., Slimani, N., Finglas, P.M., Perozzi, G., Evelo, C.T., van Ommen, B., Tetens, I., Ocké, M., Laville, M., Fezeu, L., Cavelaars, A.E., Timotijevic, L., Zimmermann, K., Poppe, K., van 't Veer, P. (submitted 2015, in revision 2016). EuroDISH: Assessing the Research, Societal, Innovation, and Policy Need for Research Infrastructures in the European Food and Health Domain, *Trends in Food Science & Technology*.
6. Snoek, H.M, Eijssen, L.M.T., Geurts, M., Vors, C., Brown, K.A., Bogaardt, M-J, Dhonukshe-Rutten, R.A.M., Evelo, C.T., Fezeu, L., Finglas, P.M., Laville, M., Ocké, M., Perozzi, G., Poppe, K., Slimani, N., Tetens, I., Timotijevic, L., Zimmermann, K., van 't Veer, P. (submitted 2016). Advancing food, nutrition and health research in Europe by connecting and building research infrastructures: Results of the EuroDISH project, *Trends in Food Science & Technology*.
7. Lusilla, P., Castellano-Tejedor, C., Helf, C., Zwickl, P., Hlavacs, H., Haukkala, A., & Ravaja, N (2012). Integrating Motivational Techniques in gamification systems for behaviour change. Poster communication presented at ICMI 2014 International Conference on Motivational Interviewing. Amsterdam, June 16-18, 2012. Poster ICMI.
8. Technology readiness levels (TRL)" (PDF). European Commission, G. Technology readiness levels (TRL), HORIZON 2020 – WORK PROGRAMME 2014-2015 General Annexes, Extract from Part 19 - Commission Decision C(2014)4995.

Measuring Food Choice and Consumption Behaviour with Real, Fake or Virtual Food Realities – a Comparative Approach from the Richfields Program

B.E. Mikkelsen[1], T. Bucher[2,3], S. Hieke[4], M.C.D. Verain[5] and J. van den Puttelaar[6]

[1]Captive Foodscape Studies; Dept. of Clinical Medicine; Aalborg University; Copenhagen, Denmark, bemi@dcm.aau.dk
[2]Institute of Food, Nutrition and Health, ETH Zurich, Zurich, Switzerland. tbucher@ehz.ch
[3]School of Health Sciences, The University of Newcastle, Newcastle NSW, Australia. tamara.bucher@newcastle.edu.au
[4]German Institute of Food Technology (DIL), Quakenbrück, Germany. s.hieke@dil-ev.de
[5]LEI Wageningen University and Research Centre, Wageningen, Netherlands. muriel.verain@wur.nl
[6]LEI Wageningen University and Research Centre, Wageningen, Netherlands. jos.vandenputtelaar@wur.nl

Abstract

Introduction: Understanding and modelling food choice is of significant interest to public health policy makers, food retailers, caterers as well as food and nutrition researchers. Behavioural nutrition and the study of pathways leading to food choice is a growing field of scientific inquiry and with recent developments in information and communications technology (ICT),) new avenues have opened for research in this field. A number of lab facilities have been set up to study behaviour and food choice. These facilities offer a range of possibilities to study food choice, purchase and consumption.
Purpose: The aim of this paper is to give an overview of selected food labs and discuss the options they present for consumer research. The paper presents how real, fake and virtual food realities can be used for food choice and behavioural nutrition experiments.
Methods: The paper analyses the strengths and weaknesses of three example food labs. It looks at real food approaches in the "Restaurant of the future" (NL) and the FoodScapeLab (DK) as well as the fake food approach at the Fake Food Buffet (CH) and cases of virtual food reality where food choice experiments can be done on-screen before carrying out an experiment with real foods.
Results: The examples of lab facilities designed to experiment with behavioural nutrition presented in this paper all offer new potentials for convenient and easy data-collection about behavioural nutrition. While settings with real food have the advantage of presenting the most familiar context for participants, it also involves higher costs and less reproducibility than when using fake foods. Regarding virtual food reality, it can lower costs even more and facilitates data collection, but its higher unfamiliarity and unknown validity need to be taken into consideration.
Conclusion: While lab based collection of data offer new avenues for studying food choice under experimental conditions the development and maintenance of such facilities is both knowledge, labour and cost intensive. Increased cooperation, knowledge sharing and research infrastructure creation would be ways to meet that challenge.

Introduction

Food choice is complex, and is influenced by numerous determinants. Studies have shown that contextual factors such as social interactions, meal duration, eating atmospherics or distractions influence food consumption [1, 2]. Besides ambient influences, cues directly related to food or the way in which it is presented can also unconsciously influence consumption volumes. In order to understand how eating and choice environments influence behavior and how they can be effectively restructured to promote healthier choices the ability to carry out experimental modelling of food choice within controlled environments is essential. Recent developments in information and communication technology (ICT) have created new possibilities for research in the field. A number of lab facilities have been set up to study behaviour and food choice. These facilities differ in the food studied and in terms of whether they focus on purchase or consumption.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

The aim of this paper is to give an overview of these recent approaches to food labs. The paper presents how real, fake and virtual food realities can be used for food choice and behavioural nutrition experiments. The paper presents an overview of the three approaches to experimenting with behavioural nutrition, with a focus on the advantages and drawbacks of each set-up and suggests directions for a future research infrastructure in this field.

Methods

The three cases cover the Fake Food Buffet at ETH Zurich (food choice), the FoodScape Lab at Aalborg University (food choice, consumption) and the Restaurant of the Future at Wageningen University (food choice, purchase and consumption). These three facilities were selected because of their geographical spread across Europe and because they cover all three behavioural aspects: food choice, purchase and consumption of food, at various levels of external validity: virtual foods - fake foods - real foods - real restaurant setting. Furthermore, they all employ different IC technologies and devices to capture and store the data. The three cases cover the FoodScapeLab at Aalborg University (food choice, consumption), the Restaurant of the Future at Wageningen University (food choice, purchase and consumption) and the Fake Food Buffet at ETH Zurich (food choice). These three facilities were selected because of their geographical spread across Europe and because they cover all three behavioural aspects: food choice, purchase and consumption of food, at various levels of external validity: virtual foods - fake foods - real foods - real restaurant setting. Furthermore, they all employ different IC technologies and devices to capture and store the data.

1. FoodScapeLab (DK)

The FoodScapeLab is a real and virtual food lab where behavioural studies can be done. It was developed for the teaching in food choice dynamics at the Integrated Food Studies [3]. The laboratory space is divided into 3 areas: COOK, SERVE and EAT, depending on the food choice focus of the research protocol. In the FoodScapeLab experiments with virtual food and real foods can be conducted. It serves as a base camp and a docking station for maintenance and calibration of the devices that are used to collect data on food choice behaviour. Conceptually, the lab is founded on the idea that it is possible to define foodscapes as a conglomerate of food, people and spaces [4], and the purpose of the lab is to make it possible to study the interactions taking place in foodscapes. Common for the data that comes out of the lab is that they all come in a very structured format. All analyses are done in the ANALYTICS section, where collected data are interfaced with background data, for instance from food composition databases or databases on carbon equivalents or ingredient prices. The devices and functionalities in the lab dealt with in this paper are related to the intelligent buffet (IB) and the foodscape tracking (FT) that both use real food and the virtual food choice simulator (VFCS) that uses virtual food reality.

Intelligent buffet. The intelligent buffet comes in a mobile (mIB) as well as in a stationary version (IB) [5,6]. The technology is designed to automatically detect food choices under experimental conditions (FoodScapeLab) or outside the lab under field conditions (living lab foodscapes). It is a traditional buffet, which has been further developed and equipped with eight scales and sensors based on RFID technology. The technology builds on the insights from the Dietary Intake Monitoring System (DIMS) that offers intelligent monitoring of food intake technology [7]. The IB operates without the digital cameras, but uses the same scales and RFID reader technology. In the IB experiment, the subjects, when enrolled, are given a RFID bracelet in order to easily register events at the buffet. The sensor detects the person via the bracelet, and the amount taken from a particular scale. The structure of the output can be seen in Table 1. Each record in the output is similar to one "event" in other words, they corresponds to every time a subject would take anything from the buffet. To sort out any disambiguation in the output, recordings from overhead cameras connected to Observer XT software can be used. In the mobile version the scales come as stand alone scales that can be placed as needed in the food environment. A typical IB experiment setup is as follows; first, the protocol for the trial is developed describing what kind of hypothesis to test. When the action has been defined and IB is set up, the test subjects are recruited. Video recording equipment is set up to record the full experiment as an extra security. The test persons are registered and enter the EAT area of the laboratory. They make their food choices, while their behaviour is recorded by IB controller software and the overhead video cameras. Apart from the requirement that people use their RFID tag before they take the food, there is nothing unusual and therefore minimum bias. When the experiment is completed, researchers and students retreat to the ANALYTICS area and analyse the data, e.g. in terms of nutrient content or climate impact.

Table 1. Intelligent Buffet data output.

Subject, no	**Time,** hour: minute: second	**Change of weight,** grams	**Scale,** no
1	00:01:55	23	1
5	00:02:59	50	8
8	00:01:39	76	7
9	00:01:49	34	6
3	00:02:51	55	5
4	00:01:50	67	5

Foodscape Tracking (FT). The foodscape tracking technology [8] follows the motion pattern of the consumer in the micro foodscapes and can be used to track for instance food choice dynamics around a buffet. It operates in an experimental mode, where sensors are set up in the foodscape, where the test persons are expected to make their choice. Instead of detecting the physical presence of consumers the FT detects the signal from the phone through the MAC address. GeoTags can then be attached to illustrate the motion pattern of the individuals studied. At the same time context sensitive questions can be asked in real time and can be compared to the answer the respondent given previously in a background questionnaire. It allows a real time ethnographic approach, where the consumer is asked in the actual behavioural situation. It can, for example, be used to investigate the difference between "saying" and "doing". In a study on secondary school students eating habits during the lunch break, we found that a third of the students who responded that they never went outside the school to eat, were actually tracked at one of the food stores or fast food places that were included in the study [8]. It should be emphasized that the smart phone version of the FT technology requires informed consent of all parties. The tracking technology comes in a slightly different, anonymous version –the Foodscape Heatmapping [5]. It is based on visual tracking of consumers using thermal cameras. This method of behavioural analysis is relatively new. It allows automatic collection of a much larger amount of data quickly, cheaply and objectively compared to manual methods. The technology has so far been used in both open spaces and indoor sports facilities [9]. The thermal technology is well suited for analyzing human food choice dynamics since the identity of the people cannot be recognized from the thermal images. In addition, the technology is independent of light, which is the weakness of many camera-based solutions. In future experiments, this tracking technology will be used to analyze consumers' movements and choices. The data that can be returned in such kind of experiments are shown in the figure grid below (Figure 1) and is about the length of time (minutes) that a given person is in a given grid of the foodscape. The figure illustrates the type of output that the Foodscape Tracker can generate. In the example the pattern of three subjects around a buffet are tracked and calculations can be made on how long each subject is in each of the squares of the foodscape. The buffet is illustrated in the middle.

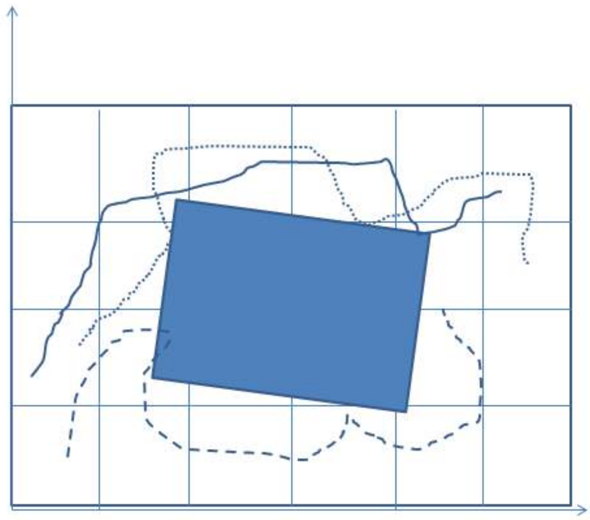


Figure 1. Foodscape Tracking data output.

C. Virtual Food Choice Simulating. As an alternative to using real foods in experiments and to save on purchasing dishwashing and food waste costs a 3D based virtual reality technology was developed under the brand name Virtual Food Choice Simulator (VFCS). The technology can be used to create a virtual food environment in which the consumer can make a virtual food choice. It can replicate buffets and supermarket without the cost of rebuilding and food. VFCS technology in the context of the Foodscape lab has been used in the study of consumer response to healthy check out aisles in supermarkets [10]. Based on questionnaires collected from customers in Lidl in a Copenhagen suburb, and their attitude to the possibility of having a healthy checkout options, a new design was developed for alternative layouts. The study found that only 10% of consumers had noticed the healthy checkout aisles. VFSC was used to develop various designs that could spice up the look. The new design was then tested in the virtual world among students and staff at the Aalborg University campus. The results showed that 83% of respondents with the new check-out design became more motivated to use the new check out aisles [11].

2. "Restaurant of the future" (NL)

The Restaurant of the Future is a real-life canteen with food lab facilities, situated on the Campus of Wageningen University and Research Centre. It consists of a buffet area with counters and a lunch area, where Wageningen UR employees and students, as well as visitors, can buy and consume their lunch [12]. For part of the participants, the Restaurant of the Future is their habitual lunch location; others are one-time visitors or go there for lunch occasionally. Daily visitors are useful in studies on dietary patterns and changes in habits due to changes in the choice context. The drawback of using daily visitors in studies with changes in the choice context is that they could become aware of these changes. Occasional or one-time visitor are interesting, when the aim is to study how they react to a specific choice context, but these one-time visitors have the bias of not showing habitual behaviour and being aware of the research context. This mix of participants is a strength of the Restaurant of the Future, although representativeness of the sample is naturally biased due to for example the large(r) number of highly educated people at a university campus.

Apart from a (semi-)natural context for the customers, the Restaurant of the Future provides a combination of opportunities for observational research and changeable surroundings. The combination of control over the surrounding, observation methods, and a population that comes in naturally makes this a distinctive research facility. The research population that enters the premises by themselves are first attended towards the possibilities of participating in a study. Upon payment, first time visitors should give a statement of informed consent, to give permission to be monitored during their lunch. Visitors that come more regularly are registered with a registration card, which makes it possible to track their food purchases over a longer period of time. In addition, this would give researchers the possibility to contact these people with additional surveys for example. The order of the buffet and the positioning of the food are changeable, as well as price labels and food information. In the lunch area, there is the possibility to change the arrangement of the tables, for example by combining small tables into one big table where many visitors can sit together. In addition, colour of the lightning, music and scent can be changed. Moreover, possibilities are present to change the infrastructure.

The unique set-up with cameras provides the opportunity to track visitors in the buffet area as well as the lunch area. The cameras can be controlled from a control room, including possibilities to zoom in on certain visitors. Images can be stored and analyzed at a later point in time. This also means that researchers or clients do not have to be present all the time at the research location when the experiments are running. Unfortunately, the analysis of camera images is time-consuming. The set-up also gives little freedom to run multiple settings simultaneously, which limits the experimental designs that are possible. This unique facility, with a lot of flexibility in the choice context, and the presence of (tracking) cameras, makes the restaurant a useful location for pre-testing of branding and communication, and product or concept acceptation (out of home). Some more concrete examples of possible research topics are:

- Food purchases (e.g. reactions on price changes or on changes on food labels).
- Tracking of the walking route by use of (tracking) cameras.
- Insights in consumption behaviour over longer periods of time.
- Impact of changes in the environment (light, sound, smell, position).

In relation to other research methods the Restaurant of the Future gives the opportunity of combining a number of techniques to study food behaviour in relation to long-term behavioural change. The Restaurant of the Future gives a unique insight by combining the possibility of habitual behaviour with cash register data per person per unique product, camera view possibilities, pathway tracking and combinations thereof.

3. Fake Food Buffet (CH)

Experiments in real life venues are difficult to control, as settings change constantly and consumer environment interactions are highly complex. Furthermore, experimental research involving real food is often limited by practical problems such as high costs, limited availability of suitable infrastructure, and the effort of preparing food. Therefore, traditional food choice experiments were often limited to very simple food selections or only single food items. The fake food buffet (FFB) is an experimental method, which can overcome these common practical limitations by using food replicas to investigate daily food choice under controlled laboratory conditions. The method allows conducting food choice experiments (e.g. product choice, portion size choice or meal composition) in under well-controlled laboratory conditions. The FFB is a buffet where consumers are invited to select meals from a range of very realistic replica foods, as used for displays [13]. The tool allows the investigation of nutrient and health claims or nutrient information on food choice, nudging effects or educational interventions as well as other manipulations under controlled laboratory conditions. In a typical FFB experiment, participants are invited to choose portions, products, meals or diets from a variety of fake food products, which appear authentic and can be portioned continuously. The foods on the buffets are carefully pre-selected and arranged by or in collaboration with nutrition experts, and they are linked with a nutrition database. Foods, portions, meals or diets selected by participants can be evaluated efficiently and are compared between experimental conditions. The method has been shown to be reliable and valid [14], and has been used in several studies, e.g. to assess how nutritional information affects consumers' meal composition in response to information [15], or to investigate whether an increase in vegetable variety is a promising strategy to improve adults' and children's food choices [16] [17]. Meanwhile, two laboratories in Europe and one in Australia have implemented the method, and two new facilities are planned in 2016. Fake food experiments are limited to food selection, as the replica foods cannot be consumed. However, the particular strengths of the method are the high controllability of various environmental cues and the cost effectiveness, which allows the investigation of food choice behaviour on complex offers.

”Food” Technology	Cases	Description of ”food tech”	Description of data capture = outcome measure	Strengths	Weaknesses
Real food	FSL	A facility with a Cook, Eat & Serve area	Observer XT, Intelligent Buffet, Heat Mapping, eTracking	Familiarity	Costs, preparation, cleaning
Real food	WUR	A canteen set-up with a (real) buffet, counters and a lunch area.	Purchase data, video of the buffet and lunch area, tracking and possibilities to measure food waste.	A real-life canteen context	Data structure, costs, single experimental condition at one moment in time
Fake food buffet	ETHZ, UoN, Konstanz (and further labs planned in 2016)	A buffet with replica food items from which subjects choose from	Portion sizes, meal composition, applied knowledge, alignment with dietary guidelines (%GDA, RDA)	low costs, no cleaning, highly controllable environment, reproducibility and validity, experiments, assessment of meal composition (complex choices) environmental influences, applied knowledge	No consumption, or food odours
Virtual food	FSL AAU	A virtual food environment that (VFCS – virtual food choice simulator) can be shaped in any style and in which consumer can shop virtually	Software/hardware based (for instance put on shopping trolley, brought to check out aisle, purchased with “virtual money”	Low cost, easy to set up experiment, no cleaning, easy and fully automatic data capture	Unfamiliarity, unknown validity

Table 2. Overview of the real, fake and virtual food experimental technologies and the food lab affiliation

Discussion

A variety of methods can be used to assess food consumption behaviour. Table 2 provides an overview of the real, fake and virtual food experimental technologies available in the three studied food labs as well as an assessment of the strengths and weaknesses. The technologies and methods can be used to answer a broad range of research questions. For instance whether taxes would be effective on changing choices or whether modelling of choice dynamics would lead to shift in choice and could be cost effective. This could in turn lead to recommendations for public health interventions. Pilots at different stages give different degrees of freedom and drawbacks. In the restaurant for example, a classical controlled experiment is not possible, which makes it difficult to run and compare multiple changes simultaneously. These drawbacks can be balanced by the use of other techniques like a virtual choice simulator [10,11] or a fake food buffet. A facility like the Restaurant of the Future, however, can shed light on changing of behaviour over time, but only in this specific context. A large part of daily consumption takes place elsewhere. To map this consumption behaviour, one needs to rely on alternative (preferably also low invasive and not self-reported) observations like smartphone apps and other trackers that people naturally take to their daily environments and activities. While many studies often have mixed results, the combination of all these techniques can shed light on the true nature of behaviour, from a fake food buffet to a real restaurant, when the data is comparable better insights into consumer behaviour can be ascertained.

With the increased need for insights in consumer food choice, there is a great need for facilities that can test choice dynamics and mechanisms underlying food choice in environments approach real life settings as close as possible. A broad range of stakeholders are potential users of such insights. And with the increased availability of intelligent devices, lab-based approaches to studying consumer choice and the role of food choice architecture is spreading, and recent developments in ICT have created new opportunities for researchers in the field of eating behaviour and food choice. We have identified and compared three of these. We conclude that they all offer new and innovative potentials. Regarding the food used, they represent the spectrum from real food to fake foods and virtual foods. Each of these labs offers strengths and weaknesses that will need to be considered when deciding on the concrete research questions and the study design. The food labs have followed slightly different directions in terms of the food studied and in terms of whether they focus on purchase or consumption. The initial insights from the study suggests that developing laboratory facilities for the study of human food choice is costly and knowledge intensive. In addition, the maintenance and operation of the facilities is costly in terms of validation, calibration and service and in all cases it requires permanent staff to manage these activities. More cooperation between the researchers and linking the facilities can make better use of the efforts stated above and make them more effective. In the future, research facilities and researchers could benefit by combining the different research methods and the data. In order to do so, standardized protocols are needed to compare, combine and link the data to finally reach more insights into food behaviour. Setting up a Research Infrastructure for sharing food data is a first step towards this future. These possible benefits should facilitated by the European strategy for food research infrastructure (ESFRI).

Acknowledgements

The present paper is one of the results from the Workpackage 8 and10 of the Richfields project. The development of the technical equipment at the Foodscapelab at AAU have been supported by the Danish Agency for Technology, Science & Technology through the dVices4food program. The Virtual Food Choice Simulating devices were developed as part of the Foosions program with the support of the AAU matchmaking program**.** The Intelligent Buffet (IB & mIB) and the Foodscape Tracker (FT) were developed in cooperation with Michal Dobroczynski, Syscore. The RICHFIELDS program is funded under the H2020 program.

References

1. Wansink, B: Environmental factors that increase the food intake and consumption volume of unknowing consumers. Annual Review of Nutrition, 24:455–79, 2004.
2. Wansink, B Mindless Eating: Why We Eat More Than We Think. Bantam Books, New York, 2007.
3. Mikkelsen, BE & Justesen, L: Multi disciplinarity in higher educations for the captive foodscape profession – case of Denmark, in symposium ”Research based educations for future foodscapes – the New NordicFood4Many Experience. In Opportunities and Challenges for Food and Eating in Society. *Proceedings of The International Conference on Culinary Arts and Sciences.* Editors Feldman, C & Murra, D. Publ. by The College of Educ. & Human Services, Montclair State University, p 262- 263.
4. Mikkelsen, B.E. (2011). Images of foodscapes. *Perspectives in Public Health*. **131**(5), 209-216
5. Mikkelsen, B.E., Dobroczynski, M.; Gade, R, Ofei, K.O.; Pawlowski, K.D. & Serafin, S. (2015). Forskere vil tracke forbrugernes adfærd [in Danish]. *Researchers track consumer behaviour*. Downloadable from www.videnskab.dk
6. Mikkelsen, B.E. & Dobroczynski, M. (2014). The potential of the Intelligent Buffet in measuring food intake in a laboratory setting. *Proceedings of Measuring Behavior 2014, Wageningen, The Netherlands, August 27-29, 2014 (Paper 60).* Editors: A.J. Spink, L.W.S. Loijens, M. Woloszynowska-Fraser & L.P.J.J. Noldus. www.measuringbehavior.org
7. Ofei, K.T., Holst, M., Rasmussen, H.H., Mikkelsen, B.E. (2015). Effect of meal portion size choice on plate waste generation among patients' with different nutritional status – An investigation using Dietary Intake Monitoring System (DIMS). *Appetite*. 91, 157-164Top of Form
8. Koirala, S. (2015) *Tracking smartphone with mobile SMS technology as dietary data collection method: A case study in Lynge School*. (Master's thesis at Integrated Food Studies) http://projekter.aau.dk/projekter/files/213622458/Final_Master_Thesis.pdf
9. Gade, R. (2015). *Taking the Temperature of Sports Arenas – Automatic Analysis of People*. PhD thesis. November 2014
10. Andersen, M., Brisson, P., Hald, P., Godtfredsen, D., Serafin, S., Mikkelsen, B.E. (2014). Validation of a Virtual Reality Tool to Test Consumer Response in Supermarket Settings. *Proceedings of Measuring Behavior 2014* (Wageningen, The Netherlands, 27-29 August 2014), p. 412-413. Editors: A.J. Spink, L.W.S. Loijens, M. Woloszynowska-Fraser & L.P.J.J. Noldus. <www.measuringbehavior.org>. Accessed 4 April 2016.
11. Cekatauskaite, V., Kwebiiha, P.F., Lindum, J., Pawlowski, K.D., Sørensen, J.R., Vejlund, C., Vilhelmsen, S., Mikkelsen, B.E. (2015). Designing a Healthy Supermarket Checkout Aisle Proofing the concept of

Virtual Food Choice Simulator. *Proceedings from ICCAS 2015* (Montclair State University, New Jersey, June, 2015), p. 108-114.

12. Van Herpen, E., Immink, V. & van den Puttelaar, J. (2016) Organics Unpacked: The Influence of Packaging on the Sales of Organic Fruits and Vegetables. Manuscript submitted for publication.
13. Replica foods. Döring GmbH, www.attrappe.de.
14. Bucher, T., van der Horst, K., Siegrist, M. (2012). The fake food buffet – a new method in nutrition behaviour research. *British Journal of Nutrition*. **107**(10), 1553-1560.
15. Bucher, T., van der Horst, K., Siegrist, M. (2013). Fruit for dessert. How people compose healthier meals. *Appetite*. **60**(1), 74-80.
16. Bucher, T., Siegrist, M., van der Horst, K. (2013). Vegetable variety: an effective strategy to increase vegetable choice in children. *Public Health Nutrition*. **17**(6), 1232-1236.
17. Bucher, T., van der Horst, K., Siegrist, M. (2011). Improvement of meal composition by vegetable variety. Public Health Nutrition. 14(8), 1357-1363.

A New Technology and Conceptual Framework for the Measurement of Human Taste

R. Kyle Palmer

Opertech Bio, Inc. Philadelphia, Pennsylvania, USA.
kpalmer@opertechbio.com

We have developed a new high throughput technology and methodology for human taste testing, called TāStation™, that departs radically from the psychophysical methods that have predominated the field of taste measurement. The TāStation™ is a fully automated sensory measurement system composed of a robotic sample delivery mechanism, a touch-sensitive display that records a subject's response to taste stimuli, and an administrative laptop computer that runs subject-interactive algorithms and coordinates the operations of the system through a database. The interactive algorithms are designed to incorporate principles of operant conditioning and signal detection theory and are structured to provide consequences to the subject's response on each trial.

A 96-well plate contains the taste stimuli and is placed on the floor of a cabinet that houses a robotic gantry that holds an automated pipette (see Figure 1). The gantry is directed by the computer program to randomly select a single well and draw a fixed volume of between 200 and 500 ul. The gantry then moves the pipette to a position within reach of the subject. The subject in turn removes the pipette from the gantry, self-administers its contents to the tongue, and makes a response on the touch-screen.

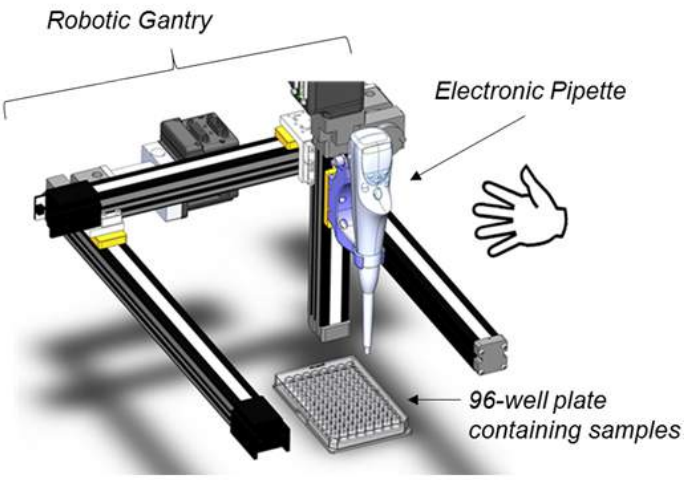


Figure 1. Robotic gantry holding an electronic pipette (cabinet housing the gantry and pipette not visible). The gantry moves the pipette over a 96-well plate to a randomly selected well and lowers the pipette into the well to draw the sample (0.2 – 0.5 ml volume). The subject manually removes the loaded pipette and self-administers the sample to the tongue.

The subject is given minimal verbal instruction on how to respond, and is instead trained through the interactive algorithm to make responses on the touch-screen that are dependent upon on his or her ability to detect and distinguish taste stimuli. The touch-screen is calibrated by designating coordinates on the screen as targets associated uniquely with specific taste stimuli. Targets can be place at any location. The entire target is defined by three concentric radii surrounding the central coordinates, similar to a dartboard. Multiple taste stimuli can be assigned to coordinates across the screen according to any desired pattern. Although a grid of any design may or may not be displayed, the targets are not made visible to the subject (see Figure 2.) Instead, subjects discover the location of the targets through the algorithms, which operate like a game, and learn to associate the targets to their assigned taste stimuli through trial and error. Hiding the targets from view avoids providing the subject with visual cues that would otherwise direct and thereby constrain the subject's responses. With fewer visual

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

cues, the subject's responses are more influenced by the behavioral processes of stimuls- and response-generalization.

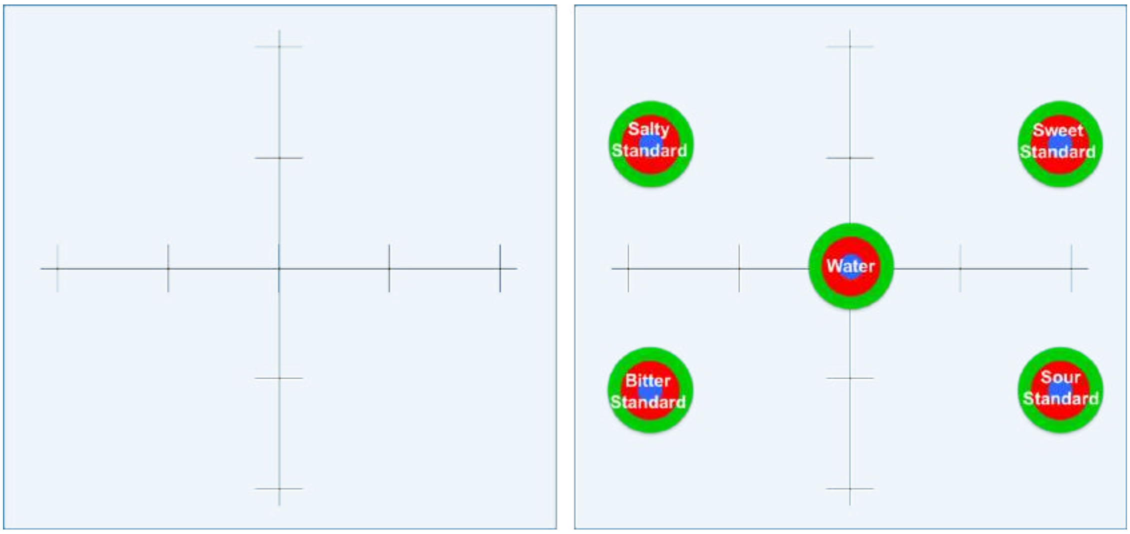


Figure 2. Schematic illustration of the touch-sensitive display on which the subject records his or her responses. The left panel represents the view of the screen presented to the subject. The right panel shows where the targets associated with the standard taste stimuli have been placed—these targets are not visible to the subject.

Touch responses that correctly match the taste stimulus to its target are rewarded by the immediate appearance of a poker chip image on the screen (see Figure 3.) Blue, red, and green virtual poker chips, representing different point values, are awarded according the accuracy of the touch response—the highest value reward results from touches made within the central radius. Touches made beyond the third, outermost radius are not occasioned by a poker chip reward, but instead result in a negative consequence, such as a pause in the game and a point reduction penalty. The cumulative point value of all poker chips earned by the end of the test session is directly remunerated with cash. Thus taste-testing performance and rate of testing are incentivized.

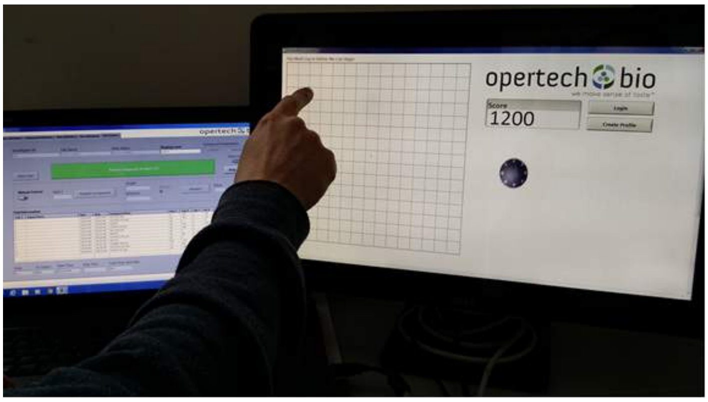


Figure 3. Subject finds the correct taste-associated target and is rewarded with a poker chip. Points earned for correct choices accumulate across trials; cumulative score also appears on the screen and is updated on each trial.

Each trial is defined by the sequence of sample presentation, sample tasting, touch-screen response, and consequence, and the duration of a trial is approximately 30 seconds. Subjects typically complete a 96-trial session in 45 minutes with high performance accuracy and test-to-test reproducibility. The software is designed to accommodate flexibility in experimental design so that a large array of test protocols can be operated under the interactive algorithms. The TāStation™ can be used to conduct rapid screening of tastant libraries, as well as perform concentration-response function characterization of multiple tastants simultaneously. Furthermore, the reward-based interactive algorithms can be used to train subjects toward progressively improved taste acuity across repeated sessions.

User Experience Affects the Development Process of Smartmeal — an Innovation for Accurate and Real-time Nutritional Information When Using Buffet Lines

T. Rautiainen and R. Tuikkanen

Department of tourism and hospitality management, Mikkeli University of Applied Sciences, Mikkeli, Finland.
teija.rautiainen@mamk.fi

Introduction

Eating behaviour is an interesting study field in public health nutrition, health promotion and consumer sciences. However, there is a lack of accurate nutritional information on individuals' food choices and dietary intake for study purposes and for guiding and supporting healthy eating and wellbeing. Individuals are also very interested in nutritional facts of their food. Technological solutions, such as smartphone applications, can assist individuals for example in self-monitoring, weight management, diet tracking or in choosing groceries and restaurants [4, 17]. These solutions have restrictions, but also potential [8, 3, 4, 14]. New technological solutions are needed to support individuals' healthy eating and research purposes [12, 1, 11].

Smartmeal is a brand new, patented and commercialized innovation which provides restaurant customers individualised real-time information on the nutritional value of their meal, for example, the amount of energy, fibre, fats, carbohydrates and protein of the portion taken. Smartmeal covers separate interfaces for customers and for kitchens to an online service, a touch screen application and a required number of Smartbuffet Units. (Figure 1). [6, 7.] The idea for this innovation arose from a research project conducted by Mikkeli University of Applied Sciences (Mamk). A company called Mealvation Oy manages the development work which has taken years and is still ongoing. This development work has also required several partners and knowledge of different fields of expertise. [16.]

Figure 1. Description of Smartmeal: There is a Smartbuffet Unit with scales and embedded software integrated into the kitchen management software on the left. Smartbuffet Unit has displays which show the name of the dish, special diets and the amount of calories in 100 grams to all customers. Customers using Smartmeal have registered and they have a remotely readable rfid card to identify themselves. The displays guide identified customers to take their portion and indicate the amount of food taken and the calories in the portion. Food items such as drinks, bread, dressing and dessert are added with a touch screen application. The pictures on the right show that registered customers can see the summary of the meals and their dietary intake for a selected period on a webpage.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

A relevant part of the technological development involved user centred design and methods that offer deeper insight into users, their interaction with products and adaption process [2, 5]. The importance of usability and ease of use were taken into consideration from the very beginning of the development process. Thus, there was a need for an exact understanding of the behaviour and actions of customers gathering meals from buffets and of the restaurant staff operating with buffet lines. Another aspect was to gather user information about the usability, acceptability and usefulness of the Smartmeal. User centred methods, naturalistic observation, video recordings, surveys and focus group interviews were used in the technical development.

Modelling the process activity to contribute to solution design

Operating principle of Smartmeal consists of a complex process of gathering and exchanging information about the identified customer, the amount of portion taken and the nutritional content of each dish. One main principle of the design was that the customers and the restaurant staff should be able to act as they usually do when using Smartbuffet Unit. Customers make several decisions at buffet lines such as where the tray is placed, if the plate is held in hand or on the tray and how the container lids are placed after peaking in or when taking a portion. [9]. Interviews, video recordings and observations of a real-life situation revealed the process and various ways of customers' actions. Interviewees found it also hard to recall the detailed process afterwards and challenging to give reasons for different actions taken or choices made during the process. Actions at buffet lines seem to be habits, automated processes which are both conscious and unconscious for users [2]. The results assisted in designing the operations model of Smartbuffet Unit.

Restaurant staff also follows a routine process when taking care of buffet lines. The process consists of turning on buffet units, providing them with food containers, spoons, lids, bread baskets, jugs plus other required food items and a decision making on the order the dishes are placed to the buffet. The food containers are filled or changed to new ones during lunch hours and the buffet units are cleaned afterwards. Same as the customers, the staff could not describe the work process in detail when interviewed. Process modelling based on interviews and supplemented with observations and video recordings were used to illustrate the process step by step. The process model brought valuable information for the development work. Comparing the process as it was and as it was expected to happen when using Smartbuffet Unit revealed that only few changes in the work process were needed.

Prototype piloting to benefit for user research and interaction

A group of volunteers (N=59) enrolled as test users for the prototype of Smartmeal in the spring 2013. The test users, staff members of Fazer Food Services, were expected to use the Smartmeal in the staff restaurant for a period of seven weeks. The test group members were asked to answer two types of questionnaires and they could also participate in focus group interviews. The study was not only used to gain more understanding of customers' perceptions but also to communicate with users. The first questionnaire was sent to the respondents twice a week for a three week period. They were asked to describe and give reasons for their emotions when using Smartmeal. The emotions were presented as drawings of nine facial expressions without a written explanation. The emotional aspect was found significant, because the user experience is not only about efficiency, but also about pleasantness and experience [10]. The second questionnaire focused more on the recognised effects and the significance of the received information during the seven weeks' test period. These themes were also the themes of focus group interviews. At each study the respondents were asked to give suggestions for improving Smartmeal from their point of view. The positive research outcome was that Smartmeal was found easy to use, it helped people to put together healthier meals and had a positive impact on their eating habits. [15.] In addition, the respondents made several justified suggestions for further development. These results also confirmed the development process was on the right track.

Present state and possibilities for research and development

Smartmeal is now in everyday use at Restaurant Kasarmina at Mamk Campus. Restaurant Kasarmina has approximately 900 daily customers and four buffet lines of which two are operating as Smartbuffets. Customers can choose their meal out of four main dishes, and a salad bar is included in each option. The customers have now a possibility to have personal, detailed nutritional information and statistics of their food. Restaurant Kasarmina, operating as a normal business based restaurant, offers a unique research environment and opportunity for Mamk researchers, lectures and students to study the aspects of user experience and customer adoption process when using a technical solution related to food and diets. One interesting research field is the everyday user experience. For example, why do the customers want to or don't want to use Smartmeal and what is needed to keep up the interest to use it constantly?

Smartmeal solves some basic problems of present reseach on eating behaviour, food choices and dietary intake [3, 17] by offering measurable and reliable facts on food taken. Smartmeal offers possibilities for different fields of research. One interesting question is to study whether the nutritional information presented simultaneously when taking a food portion has an effect on food choices: on the amounts or on the selection of dishes, and whether the possible effects are permanent. In addition to this it is significant to find out how individuals understand and adopt the received nutritional information. As Smartmeal offers new elements for promoting healthy and conscious eating, it is also interesting to study how Smartmeal can be used as a tool for nutritional guidance, for example, for university students, athletes or in nursing homes.

In addition to the nutritional database for registered customers Smartmeal provides valuable informationand statistics for the food service providers, like restaurants. The food service interface introduces, for example, minimum, maximum and average size of each dish in grams on a daily base. It can also reveal the favoured selections of dishes customers prefer to choose from Smartbuffet Units where all items are available with a fixed price. Estimating the exact amount of food to be prepared is difficult, because there is a lack of information: the number of customers and their choices are missing. The serving loss, mostly overproduction, was suggested to be the main reason for food loss in Finnish food service and restaurant outlets [13]. It is worth studying if the statistics can help in estimating the amount of food prepared and in reducing food loss.

Smartmeal offers new possibilities to study, promote and provide healthy and conscious eating. When integrated with suitable smart phone applications related to dietary intake or/and physical activity, a wider coverage of daily food consumption of an individual can be reached. At this stage Smartmeal has still a limited contribution to public health research, but as the number of Smartmeals, its users and the size of database increase, the possibilities become more comprehensive.

References

1. Azar, K.MJ., Lesser, L.I., Laing, B.Y., Stephens, J., Aurora, M.S., Burke, L.E., Palaniappan, L.P. (2013). Mobile Applications for Weight Management. American Journal of Preventive Medicine 45, 583 - 589.
2. Daae, J., Boks, C. (2015). A classification of user research methods for design for suitainable behaviour. Journal of Cleaner Production 106, 680 - 689.
3. Grandjean, A.C. (2012). Dietary intake data collection: challenges and limitations. Nutritional Reviews 70, 101 - 104.
4. Kong, F., Tan, J. (2012). DietCam: Automatic dietary assesment with mobile camera phones. Pervasive and Mobile Computing 8, 147 - 163.
5. Long, J.D., Boswell, C., Rogers, T.J., Littlefield, L.A., Estep, G., Shriver, G.E., Roman-Shriver, C., Culpepper, D., Kuenzi, G., Song, H. (2013). Effectiness of cell phones and mypyramidetracker,gov to estimate fruit and vegetable intake. Applied Nursing Research 26, 17 - 23.
6. Mealvation. Eat smarter. web-page. http://smartmeal.fi/ Access 5.2.2016.

7. Mikkelin ammattikorkeakoulu. Smart self-service lunch line provides customers with accurate nutritional value information. web-page. http://www.mamk.fi/r_d/sustainable_wellbeing/projects/smart_lunch_line Access 5.2.2016.
8. Mikkelsen, B.E., Dobroczynski. (2014). The potential of the Intelligent Buffet in measuring food intake in a laboratory setting. Proceedings of Measuring Behaviour 2014, 9th International Conference on Methods and Tecniques in Behavioural Research (Wageningen, 27 – 29 August 2014), 90 - 93.
9. Mörsky, H., Savikkomaa, M. (2010). Staff restaurant's customer's food taking process. Video research of how customers act in buffet lines [in Finnish]. Bachelor thesis. Mikkeli University of Applied Sciences, Department odf Environment and Hospitality.
10. Nurkka, P., Kujala, S., Kemppainen, K. (2009). Capturing users' perceptions of valuable experience and meaning. Journal of Engineering Design 20, 449 - 465.
11. Pagoto, S., Schneider, K., Jolic, M., DeBiasse, M., Mann, D. (2013). Evidence-Based Strategies in Weight Loss Mobile Apps. American Journal of Preventive Medicine 45, 576 - 582.
12. Sharp, D.B., Allman-Farinelli, M. (2014). Feasibility and validity of mobile phones to assess dietary intake. Nutrition 30, 1257 - 1266.
13. Silvennoinen, K., Katajuuri, J.M., Hartikainen, H., Jalkenen, L., Koivupuro, H.K., Reinikainen, A. (2012). Food waste volume and composition in the finnish supply chain: special focus on food service sector. *Proceedings of Fourth International Symposium on Energy from Biomassa and Waste* (Venice 12 – 15 November 2012).
14. Thompson, F.E., Subar, A.F., Loria, C.M., Reedy, J.L., Baranowski, T. (2010). Need for Technological Innovation in Dietary Assesment. Journal of American Dietetic Assosiation 110, 48 - 51.
15. Tuikkanen, R., Rautiainen, T. (2013). Älylinjaston käytettävyys ja käyttäjäkokemus Ravintola Huhmareessa. Working paper.
16. Tuikkanen, R., Rautiainen, T., Turkki, P. (2015). Smart buffet – an open innovation process through collaboration. In M-L. Kakkonen, M-L. Neuvonen Rauhala, eds. Researh Publication. Mikkeli University of Applied Sciences, 7 - 21.
17. Wharton, C. M., Johnston, C. S., Cunningham, B.K., Sterner, D. (2015). Dietary Self-Monitoring, But Not Dietary Quality, Improves With Use of Smartphone App Technology in an 8-Week Weight Loss Trial. Journal on Nutrition Education and Behaviour 46, 440 - 444

Using DIMS for Real-Time Monitoring of Patient Dietary Intake and Plate Waste: A Pilot Study at Herlev Hospital

K.T. Ofei[1], M.T. Dobroczynski[2] and B.E. Mikkelsen[3]

[1, 3]**Department of Clinical Medicine, Aalborg University, Copenhagen, Denmark**
[2]**SysCore ApS, Johan Krohns vej 5, 3tv 2500 Valby Denmark, [2]mic@syscore.dk**
[1]**ofeikt@dcm.aau.dk, [3]bemi@dcm.aau.dk**

The process of collecting and analyzing dietary data to monitor patient dietary intake and plate waste in hospitals is considered to be time consuming and troublesome and as a consequence often neglected in hospital wards. It is against this background that the dietary intake monitoring system (DIMS) was developed. The first prototype (DIMS 1.0) is an integrated technology based on imaging, weighing scale, infrared thermometer and RFID technology for real-time data acquisition and offline analysis of the captured data for the purpose of monitoring dietary intake and plate waste. The DIMS 1.0 has been piloted and applied in a hospital foodservice research for improved nutritional care [1, 2]. Based on the feedback from technology and as a result of the obvious potentials the DIMS 1.0 was taken into a further development. This had the goal of automatizing data analysis for real-time monitoring of dietary intake and plate waste and at the same time improves the feasibility and usability of the technology. As a result, the new DIMS 2.0 has been developed to include an integrated wireless connection and a dietary analysis software application which runs on a handheld device, mobile tablet. The integrated wireless connection allows online transmission of the data at the same time it is being captured to the mobile tablet for real-time analysis. In addition, the application offers the functionality of being used in a co-creational mode in which user inputs can be added from a mobile tablet to improve accuracy. This paper presents the development and the feasibility of using the DIMS 2.0 for real-time monitoring of patient dietary intake and plate waste in the first phase of an ongoing study aimed at evaluating a new meal serving system at Herlev hospital.

A prospective study conducted in medical and surgical wards over 9 weekdays in the pre-implementation phase. The DIMS 2.0 (see Figure 1) was used to collect paired before and after meal consumption photos and weight of plate contents for lunch and supper meal sessions. The study was approved by the hospital and all patients gave oral informed consent.

With the DIMS, we collected data on 104 meals served to patients from the food trolley, without interrupting meal serving routines. The developed system permits an investigator to assess patient meal composition, the total portion consumed and plate waste online from the mobile tablet (see Figure 2). The time frame for generating a complete patient dietary report can take up to two minutes. The interesting aspect of the DIMS 2.0 is that report on patient dietary intake and plate waste can be generated immediately after eating.

The DIMS 2.0 application facilitates a real-time monitoring of patient dietary intake and plate waste. It allows an investigator to have a quick and efficient overview of a patient with inadequate food intake. Photos of before and after meal servings may relevantly be used for guiding improved food intake in patients, and for improved communication between kitchen and clinical department.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

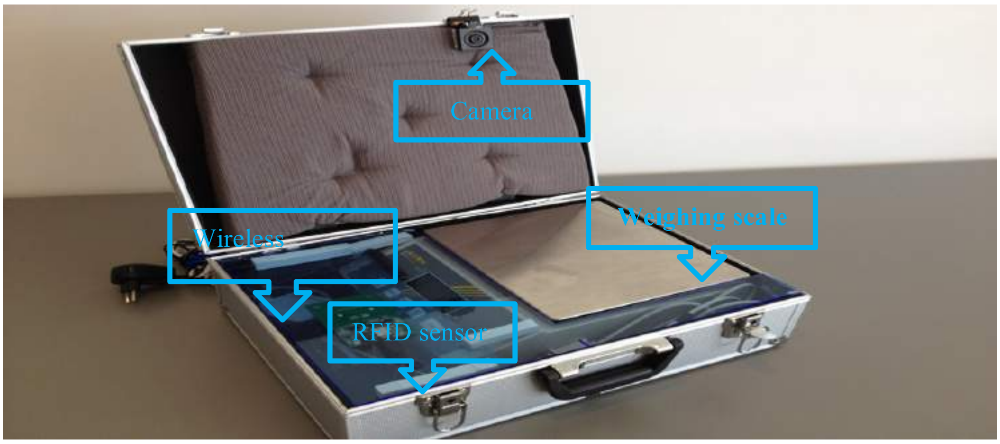


Figure 1. DIMS 2.0 for measuring before and after meal consumption photos and weight of plate contents

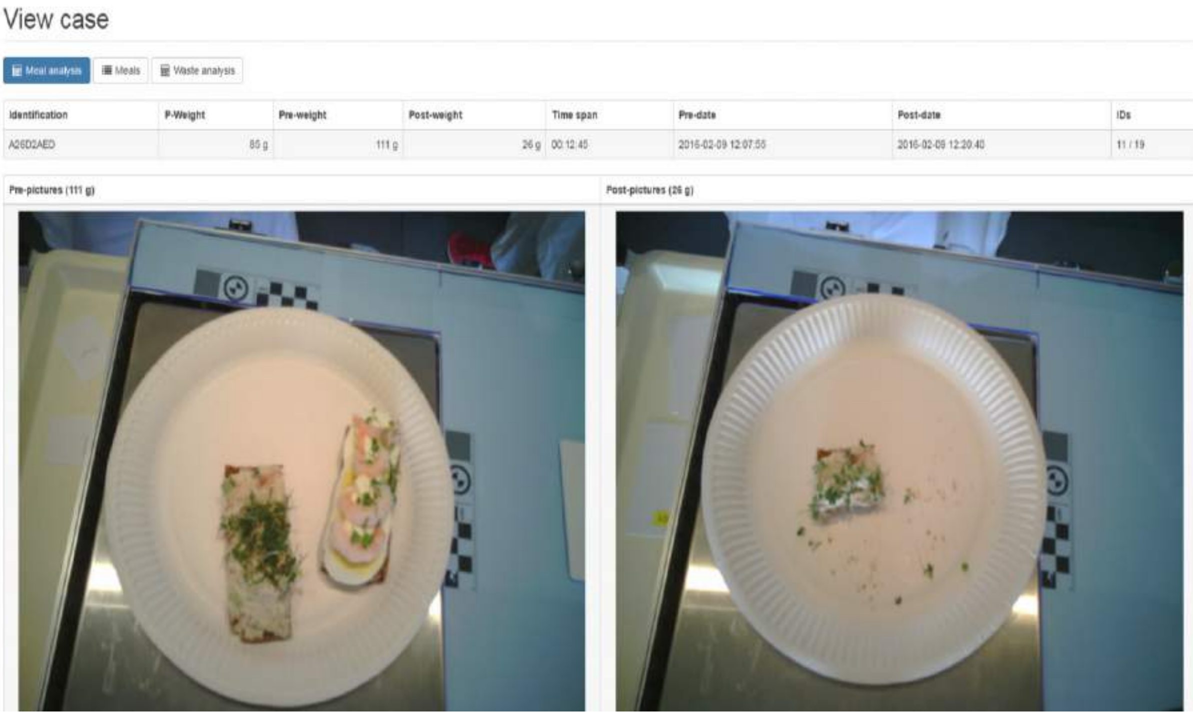


Figure 2. Screenshot of an example of the type of real-time data that can be seen on the tablet

References

1. Ofei, K. T., Dobroczynski, M., Holst, M., Rasmussen, H. H., & Mikkelsen, B. E. (2014). The Dietary Intake Monitoring System (DIMS). *Proceedings of Measuring Behavior 2014, 9th International Conference on Methods and Techniques in Behavioral Research* (Wageningen, The Netherlands, August 27-29, 2014), 94-98.
2. Ofei, K. T., Holst, M., Rasmussen, H. H., & Mikkelsen, B. E. (2015). Effect of meal portion size choice on plate waste generation among patients with different nutritional status. An investigation using Dietary 4

Sensing Behaviour using the Kinect: Identifying Characteristic Features of Instability and Poor Performance During Challenging Balancing Tasks

Daniel Leightley[1], Moi Hoon Yap[2] , Brett M. Hewitt[2] and Jamie S. McPhee[3]

[1]Centre for Military Health Research, King's College, London, England, SE5 9RJ
[2]School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, M15 6BH
[3]School of Healthcare Science, Manchester Metropolitan University, M15 6BH
dleightley@ieee.org, m.yap@mmu.ac.uk, b.hewitt@mmu.ac.uk, j.s.mcphee@mmu.ac.uk

Abstract

A framework is proposed to utilise the Microsoft Kinect One, a low-cost, unobtrusive and reliable sensing device to analyse the apprehension that exists in human motion to aid clinical decision making. We extract kinematic features, Centre-of-Mass and Body Movement Zone to determine the spatial-temporal directional changes for balancing tasks. This framework includes a pipeline that obtains the Kinect skeletal data and applies techniques to interpret the information to determine the steadiness or instability, hence providing an objective indicator of the difficulty the participant has completing the task. We conclude with a summary of our initial findings, and highlight areas for further research.

Introduction

In recent years, there has been an interest in digitalised methods for detection, analysis and quantification of human motion. This is due to the increased availability of low-cost multi-modality marker-less capturing devices [1]–[3]. The release of the latest edition in the Microsoft Kinect (henceforth Kinect) series, Kinect One, has enabled a new and improved immersive gaming experience. Kinect can be utilised for additional applications other than gaming or entertainment, most notably within the medical domain. Bigy *et al*. [4] proposed a technique for recognising posture and Freezing of Gait in those with Parkinson's disease to aid in detecting trips and falls within the home. Leightley *et al*. [5] introduced a classification framework to detect motions commonly found in a rehabilitation setting to support patient wellbeing. Yang *et al*. [6] implemented a framework that extracts both depth and colour image data from the Kinect to assess the posture of participants when performing standing balance, the framework allowed for detection of subtle directional changes such as postural sway.

The Kinect series has been utilised in a range of medical applications, to support this several studies have sought to validate the Kinect and its viability in the medical domain. Clark *et al*. [7] (revisited in [8]) captured a large number of participants performing a series of balance motions consisting of single and double limb support. Kinect and marker-based Vicon data were captured concurrently, with data from both systems filtered and synchronised. The Kinect was found to be highly robust and accurate when compared to the '*gold-standard*' of the Vicon. Mentiplay *et al*. [9] assessed the validity of the Kinect in tracking gait and its inter-day reliability when compared to 3DMA marker-based camera system. The authors found that while the Kinect is not suitable for tracking lower body kinematic data, it is capable of measuring spatiotemporal aspects of gait. Gonzalez *et al*. [10] implemented a framework that combined the Kinect and Wii Balance board to extract Centre-of-Mass (CoM), the authors found a strong correlation between the two devices and were able to assess standing and walking motions amongst a small population sample. These works detail the validity and clinical feasibility of the Kinect to assess kinematic strategies related to gait and posture.

In this work, we propose a new framework for *sensing the behaviour* of motion with the Kinect to identify difficulty people have performing balancing tasks. Progressing from previous studies [3], [11], we utilise CoM and Body Movement Zone (BMZ), coupled with the temporal domain to provide a low latency outcome for potential utilisation in *real world* situations. This will provide greater insight for practitioners and allow us to

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

identify when a participant is unsteady, in distress or finding the motion troublesome to perform/execute. This would allow the practitioner to implement coping strategies or develop interventions.

Sensing behaviour with Kinect

Several works have been proposed to measure and analyse behaviour present in human motion. The most prevalent methodology suggests using one or a combination of intrusive sensors, such as body-based accelerometer or reflective-markers. Then, the clinically relevant indicators are extracted by analysing the patterns presented in time series data generated by these data mediums [12]. In recent years, the Computer Science community has proposed an array of solutions. These works have predominantly focused on depth sensor technology, which has been shown sufficiently accurate and responsive for tracking in both in-home and medical settings [5], [10].

These frameworks follow a similar structure. They first seek to identify the human motion, using motion classification techniques, and then undertake quantitative analysis and feature extraction techniques on the motion to provide greater understanding. Depth-sensors have been used for the assessment of balance by extracting gait-based features such as CoM from a skeletal stream to provide mobility indicators (*e.g.* [6], [8], [13]–[16]). For example, Dolatabadi *et al*. [14] proposed a home-based system for assessing changes in gait and balance using kinematic features such as CoM. The authors utilise a Microsoft Kinect sensor to observe gait recovery in a participant that had undergone surgery. They were able to track the gait changes over a number of weeks, helping to inform clinical judgements based on the information extracted. Typical gait patterns require stability standing on one-leg which is a task that becomes increasingly difficult for people with movement impairments, yet very little research has focussed on sensing stability.

The aim of the present study was to develop automated methods to sense the instability in human motion using depth sensor technology, CoM and body-movement-zone including a wide range of participant characteristics. Our work relies on several key computer vision and pattern recognition techniques such as stereovision, pose estimation, and feature extraction/representation.

Identifying Behaviour and Kinect: Framework

An overarching framework for analysing human motion obtained from the Kinect to identify unease in motion is proposed (see Figure 1). This section will introduce the framework methodology, utilisation and implementation for possible *real world* deployment. The analysis framework has been implemented in Matlab 2014a and all source code is openly available at (Link upon acceptance).

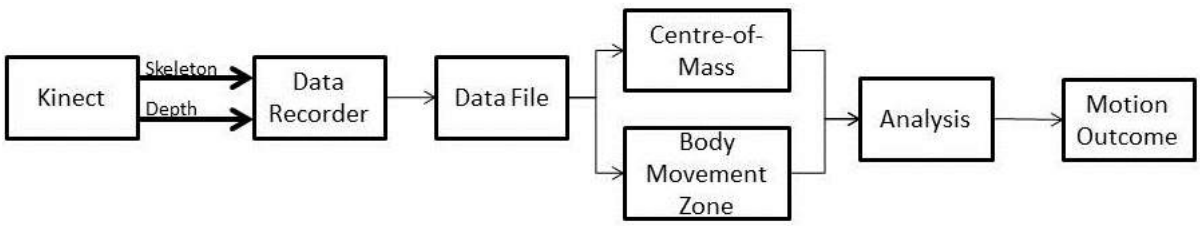


Figure 1 - Overarching framework for analysing human motion collected from the Microsoft Kinect to identify movement hesitation.

Variables	Young (≤59yrs, n=5)	Old (≥60yrs, n=5)	Athletic Old (≥60yrs, n=5)
Age	25.6 (7.03)	73.4 (3.36)	64.8 (6.34)
Height	178.6 (8.26)	172.8 (4.54)	172.8 (4.54)
Weight	75.2 (12.07)	79.4 (21.04)	62 (10.93)
Sex (male %)	70%	40%	60%
BMI	20.7 (3.63)	26.46 (7.12)	21.9 (1.56)

Table 1: Participant demographics used in this study, extracted from the K3Da Dataset [17].

In this work, the newly released K3Da Dataset [17] was employed. The dataset consists of more than 50 participants performing a range of clinically relevant movements captured by a Kinect One. The dataset is composed of three participant groups, young (≤59yrs), old (≥60yrs) and athletic old (≥60yrs). In this work, fifteen participants were randomly selected (see Table 1 for demographic information), five from each group with the following motions extracted: One-Leg Balance (Eyes Open) and One-Leg Balance (Eyes Closed), providing a total of 52 motions. Each trial was independently assessed by two coders (using skeletal and video streams) and categorised as follows: "Stable and successful", "unstable and successful" and "unstable and unsuccessful", providing ground truth information to validate the proposed framework. The following definitions were used when determining category:

- *Stable and successful*: The participant had very little body movement and was able to complete the balancing task.
- *Unstable and successful*: The participant had high amounts of extraneous body movement, but was able to complete the balancing task.
- *Unstable and unsuccessful*: The participant had high amounts of extraneous body movement and was unable to complete the balancing task.

The first coding iteration resulted in an inter-reliability of 0.84. In the second iteration, the coders consulted one-another to reach final agreement on the rating of trials. The final categorisation composition can be found in **Table 2**.

Categorisation	Young	Old	Athletic Old
Stable and successful	9	4	6
Unstable and successful	5	2	3
Unstable and unsuccessful	1	13	9

Table 2: Categorisation labelling for each participant group.

To identify the level of stability, each motion sequence was analysed **independently**, hence only the motion itself is evaluated and not dependent on any other motion that the participant (or others) had performed. While the Kinect provides a skeletal stream, similar to motion capture from a marker-based system (such as Vicon), it does so without the need for placing markers on the participants body. The Kinect utilises depth sensor technology to identify key anatomical landmarks on the body that are tracked over time (up to 30Hz). In this work, we decompose the Kinect skeletal stream into two time-series parameters, CoM and BMZ to describe the participants' behaviour at each time period (frame). This enables an overall determination of the behaviour and performance of the participant.

Using three joint locations provided by the Kinect (centre of the hip, left shoulder and right shoulder), while on their own they can provide descriptive information about the motion we seek to derive a single variable that is powerful enough to describe the motion. The CoM [10], [13] is encoded for each frame as the average point between the aforementioned joints. The change in position between consecutive frames is the directional change over time. Therefore, for each period of time we are able to identify the position, and the temporal directional movement of the CoM.

In this work, we introduce the BMZ, a parameter that describes the total volume of space occupied by the participant when in motion (see Figure 2 for a visual example). This is computed by identifying the total space covered (or occupied) by the participant's skeleton when undertaking a motion per frame using standard volume calculations. For example, if the participant is stable, with little movement, the BMZ variable is small, whereas with large variations in motion such as raising of the arm the size of the overtime BMZ increases.

To determine the behaviour exhibited by the participant, the CoM and BMZ are firstly assessed individually, and then assessed together to determine an outcome. For each parameter, the infra-framed variation is computed (*e.g.* difference between frame 1 and frame 2), using this we are able to generate the percentage difference between the current and last frame. Hence, we can represent the CoM and BMZ parameters with a percentage change value over time (*e.g.* if we have 300 frames, for each parameter we will have 300 percentage values). Therefore, we are able to identify sudden and large motion variations that are typically present when the participant has encountered difficult when undertaking a motion. Each frame (represented by the percentage value for both CoM and BMZ) is categorised as follows, based on our experience working in the field and other similar works, if the percentage is:

1. Less than 30% it is labelled as ***Stable and successful***.
2. Between 31% and 70% labelled as ***Unstable and successful***.
3. Greater than 71% labelled as ***Unstable and unsuccessful***.

Having represented each parameter by a set of labels representing the state of the motion at a specific time period, these are aggregated; the category with the most "votes" is identified as the classification for the motion.

Figure 2 - Example of a skeleton within a Body Movement Zone. Defined as the space occupied by the participant for each frame. Observe the entire sequence (represent by blue) and the surrounding box (represented by black), which is the total space occupied by the participant.

Identifying Behaviour

The ground truth labels derived in this work are used to evaluate the proposed framework for the task of detecting unease in motion performance across a range of participants. For evaluation, random samples of participants were identified, representing a large range of demographics profiles. The results were repeated five times to ensure reproducibility in the coding – no anomalies were identified. Each motion was assessed individually and compared to the ground truth labelling to determine the success of the framework.

	Young (accuracy)	Old (accuracy)	Athletic Old (accuracy)
Stable and successful	9 (0.96)	4 (0.89)	6 (0.93)
Unstable and successful	5 (0.97)	2 (0.91)	3 (0.94)
Unstable and unsuccessful	1 (0.90)	13 (0.93)	9 (0.92)
Average:	**0.94**	**0.91**	**0.93**

Table 3: Classification accuracy for each participant group based on categorisation.

The framework provided a high degree of accuracy when identifying mobility of participants who had been identified as finding the motion difficult during ground truth labelling (see **Table 3**). Overall, the framework performed robustly with a classification accuracy of 0.92, which is considerably high for the type of data being classified. Confusion observed across the participant range low, with a high rate of true positives, and a small rate of false positives and a robust overall sensitivity of 0.94, specificity of 0.97 and Matthew Correlation Coefficient score of 0.97. There was no significant difference between One Leg Balance (Eyes Open) and One Leg Balance (Eyes Closed), with an overall accuracy of 0.95 and 0.95 respectively.

Discussion and conclusion

There is little doubt to the benefits of utilising the Kinect sensor for use in monitoring, quantify and evaluating human motion. This work has relied upon the K3Da dataset, which has provided a large number of clinically relevant motions and a dives range of participants, however it is not without its challenges. First, range of the sensor is limited – only capering motions within a 4 metre by 4 metre areas. Second, Joint occlusion and sensor accuracy varied for each participant, and the types of motion being undertaken. However, when the motions were coded the coders had access to the video stream and not the skeleton alone – allowing for greater interpretation. Conversely, the framework only utilised skeletal information yet was able to identify motion stability with a high degree of accuracy.

In this work, we have sought to propose a framework for identifying instability in motion, allowing for a practitioner to intervene and identify what may lie behind the instability. Only a single outcome is provided which is able to provide an overview of what is being observed. However, it would be more helpful to identify specific temporal dynamics to improve the clinical decision making process. Further, we have utilised two motion groups, One-Leg Balance (Eyes Open) and One-Leg Balance (Eyes Closed). Future work will explore the possibility of adapting the framework to provide detailed joint-level temporal kinematic insight; hence practitioners can focus their efforts in improving confidence in these areas.

References

1. Y.-J. Chang, S.-F. Chen, and J.-D. Huang, “A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities,” *Res. Dev. Disabil.*, vol. 32, no. 6, pp. 2566–2570, 2011.
2. A. Ejupi, M. Brodie, Y. J. Gschwind, S. R. Lord, W. L. Zagler, and K. Delbaere, “Kinect-Based Five-Times-Sit-to-Stand Test for Clinical and In-Home Assessment of Fall Risk in Older People,” *J. Gerontol.*, 2015.
3. A. Kargar, A. Mollahosseini, T. Struemph, W. Pace, R. Nielsen, and M. Mahoor, “Automatic measurement of physical mobility in Get-Up-and-Go Test using kinect sensor,” in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014.
4. A. Amini Maghsoud Bigy, K. Banitsas, A. Badii, and J. Cosmas, “Recognition of postures and freezing of gait in Parkinson’s disease patients using Microsoft Kinect Sensor,” in *IEEE Conference on Neural Engineering*, 2015, pp. 731–734.
5. D. Leightley, J. Darby, B. Li, J. S. Mcphee, and M. H. Yap, “Human Activity Recognition for Physical Rehabilitation,” in *International Conference on Systems, Man and Cybernetics*, 2013.
6. Y. Yang, F. Pu, Y. Li, S. Li, Y. Fan, and D. Li, “Reliability and Validity of Kinect RGB-D Sensor for Assessing Standing Balance,” *Sensors J.*, pp. 1633–1638, 2014.

7. R. Clark, Y.-H. Pua, K. Fortin, C. Ritchie, K. Webster, L. Denehy, and A. Bryant, "Validity of the Microsoft Kinect for assessment of postural control," *Gait Posture*, vol. 36, no. 3, pp. 372–377, Jul. 2012.
8. R. Clark, Y.-H. Pua, C. Oliveira, K. Bower, S. Thilarajah, R. McGaw, K. Hasanki, and B. Mentiplay, "Reliability and concurrent validity of the Microsoft Kinect V2 for assessment of standing balance and postural control," *Gait Posture*, 2015.
9. B. F. Mentiplay, L. G. Perraton, K. J. Bower, Y. H. Pua, R. McGaw, S. Heywood, and R. A. Clark, "Gait assessment using the Microsoft Xbox One Kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables," *J. Biomech.*, vol. 48, no. 10, pp. 2166–2170, 2015.
10. A. Gonzalez, M. Hayashibe, and P. Fraisse, "Estimation of the center of mass with Kinect and Wii balance board," in *International Conference on Intelligent Robots and Systems*, 2012, pp. 1023–1028.
11. W. van Teijlingen, E. L. van den Broek, R. Konemann, and J. G. M. Schavemaker, "Towards Sensing Behavior Using the Kinect," in *Proceedings of Measuring Behavior*, 2012.
12. E. Shmueli, V. K. Singh, B. Lepri, and A. Pentland, "Sensing, Understanding, and Shaping Social Behavior," *IEEE Trans. Comput. Soc. Syst.*, vol. 1, no. 1, pp. 22–34, 2014.
13. B. Lange, C.-Y. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas, "Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft kinect sensor," in *International Conference of Engineering in Medicine and Biology Society*, 2011, pp. 1831–1834.
14. E. Dolatabadi, B. Taati, G. S. Parra-Dominguez, and A. Mihailidis, "A Markerless Motion Tracking Approach To Understand Changes In Gait And Balance: A Case Study," in *ehabilitation Engineering and Assistive Technology Society of North America*, 2013.
15. N. Vernadakis, V. Derri, E. Tsitskari, and P. Antoniou, "The effect of Xbox Kinect intervention on balance ability for previously injured young competitive male athletes: A preliminary study," *Phys. Ther. Sport*, vol. 15, no. 3, pp. 148–155, 2014.
16. O. Celiktutan, A. C. Burak, C. Wolf, and B. Sankur, "Graph-Based Analysis of Physical Exercise Actions," in *ACM Multimedia Workshop on Multimedia Indexing and Information Retrieval for Healthcare*, 2013.
17. D. Leightley, M. H. Yap, Y. B. J. Coulson, and J. S. Mcphee, "Benchmarking Human Motion Analysis Using Kinect One: an open source dataset," in *IEEE Conference of sia-Pacific Signal and Information Processing Association*, 2015.

Optimization of Automated Health Programs by Simulating User Behaviors and Program Effects

Arlette van Wissen, Aki Härmä, Rim Helaoui, Illapha Cuba Gyllensten, and Dietwig Lowet

Philips Research, Eindhoven, the Netherlands. arlette.van.wissen@philips.com

Abstract

Many health programs aim to encourage and support healthy behaviors, such as increasing one's physical activity and improving one's dietary patterns. Tailoring such programs to the constantly changing user state, is a challenge most programs struggle with. Conventionally, this tailoring has been the job of healthcare professionals and coaches who use their experience and knowledge in discussions with patients in order to understand the patient's needs and desires. In online or automated coaching programs, data-analytics based on sensor data and user input could provide an alternative method to provide tailored experiences, for instance by using behavior simulations and predictions to match the support to the (predicted) user state. In this paper we describe a method to optimize a health program for a patient based on holistic simulations based on both the behavior and physiology of the patient. The optimization target is the simultaneous maximization of the health benefits and minimization of the burden of the proposed lifestyle interventions for the patient. In this work we demonstrate how the burden of interventions can be modeled using a specific lifestyle cost metric, the Personal Cost Unit. We show how users' lifestyle behavior and physiological processes can be simulated in order to obtain a prediction for both target outcomes, given the content and structure of a health program.

Approach

It has been demonstrated that lifestyle change can have a huge impact on the prognosis of and recovery from conditions such as cardiovascular diseases, diabetes, or obesity. Yet changing lifestyle behaviors is a complex process that depends on many factors, of which personal habits, personal experience, mental state, and social, economic, environmental factors, are only a few. As such, people could use a little help when it comes to behavior change. Nowadays, information technology can provide guidance and support in this area, one can think for instance of smartphone apps that provide health programs based on measured data. These automated programs can support personal health self-management by offering different modules that target different aspects of behavior (change) or different domains. The modules that aim at increasing physical activity may for example recommend users to take more steps at work in order to increase physical activity and lose some weight.

According to Payne and colleagues, the latest count identifies more than 31.000 health related mobile apps that are currently available [7]. Most of these apps incorporate at least one prominent health behavior theory-based strategy, of which self-monitoring, cues to action and feedback are among the most [7]. According to the goal setting theory [10], the more ambitious and specific the proposed goals are formulated, the more improvement in behavior can be expected. For instance, research show that creating implementation intentions, which predetermine the when the where and the how aspects of the goal-directed behavior, helps to achieve goal attainment [6]. However, in most online health apps health behavior goals are still defined on a relatively high abstraction level, e.g. do 10% more physical activity, increase number of steps per week with 2000 steps, or eat 300 kcal less per day. These type of goals lack specificity and therefore are likely not to lead to optimal improvement. An opportunity for automated coaching systems is to select and propose more concrete goals. However, the more concrete a goal is specified (e.g. take a lunch walk today, go running on Sunday afternoon for 30 minutes), the higher the risk that the proposed goal does not fit the lifestyle and preferences of the user.

The challenge arises in how to match the support given by the program on how to accomplish this goal to the user's preferences and state. That is, if the user finds the burden of implementing the change high, the chance of successful implementation of this change becomes low. For example, the suggestion to stand up and walk around can be unpractical in a work environment with limited opportunities to take a stroll or when many meetings occur. It could be that the user can more easily incorporate walks at home, for example, or during a commute. In addition, in many offered programs the long-term outcomes of the proposed health interventions are unclear or hard to foresee.

To overcome these challenges, a health program should be able to simulate the expected impact of the interventions it provides. Based on this expected impact, the program can give the user content or modules that lead to a maximally effective change while minimizing the burden that executing this change has for the user. Although this is somewhat addressed in conventional health programs in live discussions with human coaches, the authors are not aware of any previous approaches of a structured method to achieve these two targets in a fully automated health program. In this paper we focus on an example of a health program aiming at maximizing the expected weight loss through increasing physical activity such that the program's interventions result in a minimal implementation burden for the user.

Segmenting Lifestyle Patterns

A large part of the possible burden a user might experience in implementing the suggested change, relates to the social, economic, and environmental context that determines what a person does and is able to do. For example, the user has to go to work and perform various other duties of daily living, which limits the choices and the time available for activities proposed by a health program. This is reflected in the concept of *health lifestyle*, which is in literature often defined as collective patterns of health-related behavior based on choices from options available to people according to their life chances [2][3]. In addition to context, psychological factors, such as self-efficacy, also play an important role in a person's ability to adopt new health behaviors (e.g. [11]) but those are often difficult to measure and especially to predict.

When considering the intervention burden for a user, we propose to use the user's high-level lifestyle patterns as a starting point. This structure identifies the temperal pattern that people generally adhere to. It can often be assumed relatively stable over time and will be used as a contextual framework for the health program optimization. We then model daily lifestyle behavior as a segmental pattern where the burden is uniform in each segment. This simplifies the optimization problem and makes it easier to communicate the final program to the user.

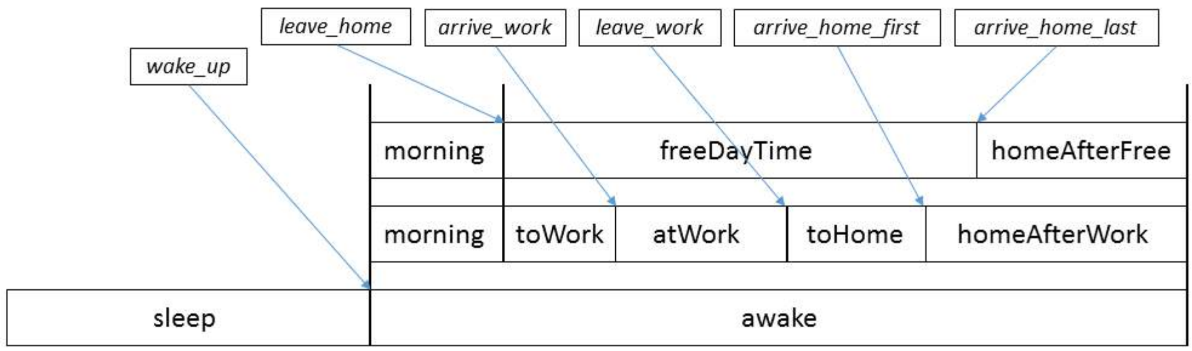


Figure 1 Segmental lifestyle model of an office worker.

In this paper we focus on the lifestyle patterns of a typical office worker. A semantic temporal model of a lifestyle pattern is illustrated in Figure 1[1]. For estimation of this model, typically one can use well-known change-point detection and template matching techniques to fit the model to data (see e.g., [4][12-14]). First, a day is divided into sleep and awake segments based on the detection of the main sleep period of the day. The

[1] The validation of the model and estimation of the parameters for an individual user are out of scope of this paper.

awake segment is further divided into two sequences of segments depending on whether a day is a work day (bottom) or a day when the subject was not at work. In the case of a workday the segments contain commute and working hours. In the case of a free day the awake segment is divided in three parts based on the first exit from home and last return to the home in the awake segment.

Personal Cost Units

In order to minimize the implementation burden for the user we introduce a segmental cost measure, called the Personal Cost Unit (PCU). The PCU can be considered an indicator of the match between the user's preferences and the proposed intervention. Hence, the higher the PCU, the less attractive it is for the user to follow-up on this intervention. In case of a very low PCU, we expect the user to be able to execute a proposed change without much difficulty. The PCU is defined for each intervention and context. Initially, they can be defined for groups of users with similar profiles, e.g. mothers with a day job or elderly who are retired. For example, the expected PCU for additional walking in a work context is be predicted to be higher than the PCU for that same intervention at home in weekends. Ideally however, the PCUs are learned for each user individually, enabling a truly tailored adaptation of the program. In such cases, the values for PCUs can be inferred from measurements from smart devices, such as the user's location, calendar, or sleep/wake patterns.

In order to model and predict the weight loss of a user we have implemented a published macroscopic metabolism model [9]. This model estimates the likely changes in weight resulting from changes in activity or dietary behavior based on initialized constants of resting metabolic rate, physical activity and body composition (identified from general demographic data). These constants can be tuned to recordings of continuous behavioral data during a baseline period, in order to improve and update the predictions. With this model, realistic changes to the weight given different interventions can be simulated assuming that the user adheres to the intervention. For example, the effect of a user's incremental change of climbing the stairs at work instead of taking the elevator can be compared to weekend runs or changes in dietary patterns. This will affect the interventions that will be judged most promising by the system with respect to optimizing the balance between intended effect and personal cost. Furthermore, during the program the continuous user measurements can be compared to the program prescribed activity levels and the predicted weight changes. Issues in adherence to activity levels and any compensatory energy intake behavior can thus be inferred [8], which could be used to provide additional coaching to help the user reach their target goal or to re-assess the goal and the personal costs of the user.

Simulating Behaviors and Effects

To select the optimal program that supports the user in reaching a specific goal, the PCUs of different options are minimized over the constraint of meeting the personal goal. One possible way to solve this is to use linear programming. Linear programming (ILP) [5] is a technique for the optimization of a linear objective function, subject to linear equality and linear inequality constraints. Figure 2 depicts an example of how ILP can be applied to a typical routine of an office worker. Given his/her daily number of steps (6000 in the example), the target of the optimization would be to find the best way to schedule additional steps for the user so that they reach their daily target of 10 000 steps. The best schedule would be the one that minimizes the personal costs (PCUs) of the suggested interventions (i.e. the additional steps) while respecting the model's constraints such as the available time in the different day segments. According to the illustrated example, the user would be more reluctant to adopt interventions during commuting to work than after work. This can be depicted in the higher cost attributed to the first commuting segment compared to the after work ones. Based on these PCUs three possible schedules are visualized with the corresponding total costs. The first is considered an invalid suggestion since no interventions are allowed during sleeping. The second is a feasible, yet very expensive solution, where the user is invited to increase their daily steps during commuting to work and during their working time. Finally, the optimal solution is represented. It consists in adding 1000 steps during work, 1000 steps during commuting to home and 2000 steps after they arrive home. Such a program would yield the minimal cost (2900 PCUs) given the model's constraints.

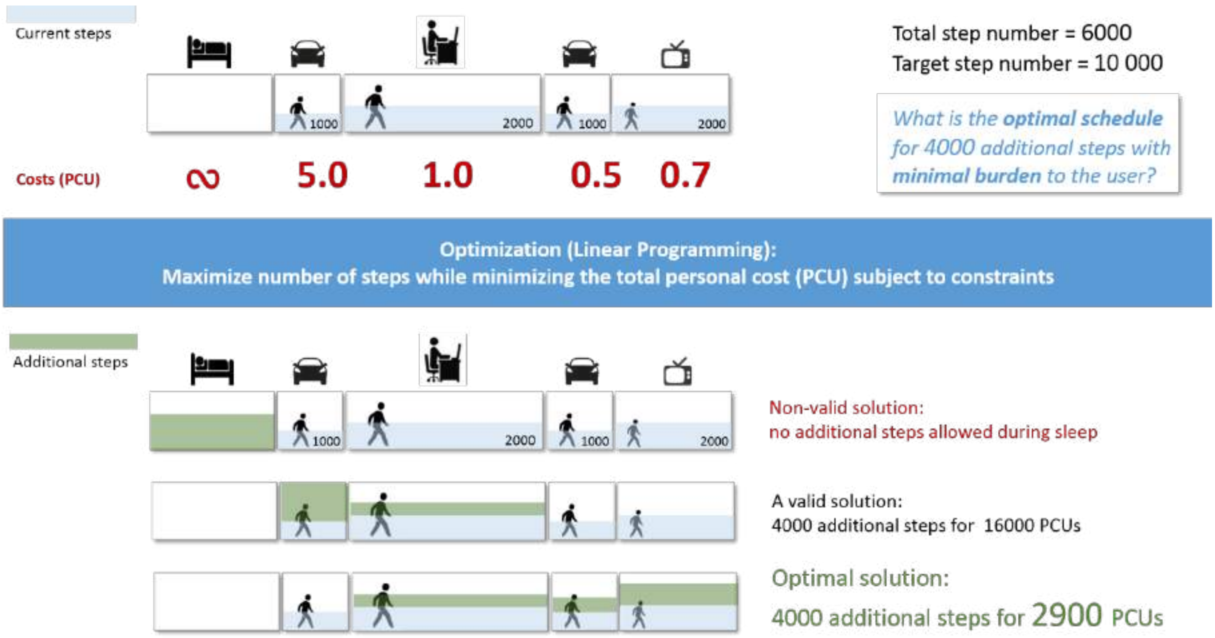


Figure 2: An example of applying ILP to optimize the intervention plan to reach a daily target of 10000 steps.

How these interventions link to expected weight loss is illustrated in Figure 3 and Figure 4. Here a simulation is shown of the routines for an office worker. In the upper bar graphs, the number of steps of a user are simulated for different segments of the day (e.g. at work, in free time in the weekend). Based on a baseline model of the user (learned from data), the personal costs are simulated for each intervention. It is important to note that the costs mean something different to each user, and their value can be used in comparative analysis. Figure 3 shows that by walking 30 additional minutes at work each day, this user can reach a 0.6 kg decrease in body weight, yet at a personal cost of 7200. Figure 4 shows the same user but instead of walking 30 minutes more each day, he walks for 2 hours every day of the weekend. This results in a weight loss of 1.2 kg, at a cost of 864. So in this program, the user will lose more weight at less personal cost. These simulations show that even a slight modification in weekly activity routines may lead to non-trivial changes in the target outcomes of a user's weight and accumulated PCUs.

Until now we discussed only examples of obtaining the value of PCUs using measurable data collected unobtrusively by smart devices. However, in order to get accurate estimations of the PCUs for each user in different contexts, different methods of establishing their values have to be considered. For example, users can be asked to give feedback as soon as they receive the intervention, or, alternatively, users can be asked to imagine different scenarios and interventions and to rate interventions according to their perceived costs. Importantly, also

Figure 3. Simulated weight change and PCUs for a user who walks an additional 30 mins each day.

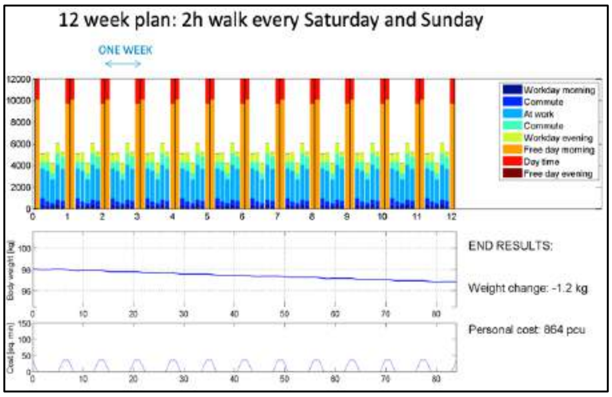


Figure 4. Simulated weight change and PCUs for a user who walks an additional 2 hours each weekend day.

psychological and social costs could be taken into account. Surveys can be used to identify these opportunities and preferences for people with different profiles. For example, in a small pilot survey (N=11), we asked office workers to indicate their opportunity and preference for different types of interventions during the segments as shown in Figure 1. Preliminary results show that the majority of people have the least opportunity to act on physical activity interventions on week days (Monday-Friday) during the day, and the most opportunity during weekdays at night and during weekend days during the day. Additionally, most people prefer the interventions during weekend days during the day. These results can be used to identify more accurately the PCUs for people during those segments, and are in line with the current assumptions that were made in the model.

Discussion & Conclusion

The paper discusses an approach to tailor automated coaching programs to the user, taking into account the predicted effects of the proposed behavior change and of costs for the user to act on them. In this paper we show how a macroscopic metabolic model can be implemented and used in simulations to predict effects of different proposed interventions. To calculate the Personal Cost Units for a user for these interventions, unobtrusive measurements about behavior and segmented context patterns can be used. However, as validation of the PCUs is important for this approach, in the future more factors should be taken into account and measured (using for example surveys) in order to gain insights into people's assigned costs. PCUs can aid the program in providing the user with interventions that lead to a maximally effective change while minimizing the burden that executing this change has for the user.

References

1. Barber, D., & Cemgil, A. T. (2010). Graphical models for time-series. *Signal Processing Magazine, IEEE* 27, 18-28.
2. Cockerham, W. C. (2005). Health lifestyle theory and the convergence of agency and structure. *Journal of health and social behavior* 46, 51-67.
3. Cockerham, W. C., Rütten, A., & Abel, T. (1997). Conceptualizing contemporary health lifestyles. *The Sociological Quarterly* 38, 321-342.
4. Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, 233-244.
5. Gass, S. I. (1958). Linear programming. *John Wiley & Sons*, Inc.
6. Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology* 38, 69-119
7. Payne, H. E., Lister, C., West, J. H., & Bernhardt, J. M. (2015). Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. *JMIR mHealth and uHealth* 3.

8. Hall, K. D., & Chow, C. C. (2011). Estimating changes in free-living energy intake and its confidence interval. *The American journal of clinical nutrition* 94, 66-74.
9. Hall, K. D., Sacks, G., Chandramohan, D., Chow, C. C., Wang, Y. C., Gortmaker, S. L., & Swinburn, B. A. (2011). Quantification of the effect of energy imbalance on bodyweight. *The Lancet* 378, 826-837.
10. Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current directions in psychological science* 15, 265-268.
11. Luszczynska, A., & Schwarzer, R. (2005). Social cognitive theory. *In M. Conner & P. Norman (Eds.), Predicting health behaviour* (2nd ed. rev., pp. 127–169). Buckingham, England: Open University Press.
12. Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers* 32, 93-110.
13. Panagiotou, V. (2015). *Blind segmentation of time-series: A two-level approach* (Doctoral dissertation, TU Delft, Delft University of Technology).
14. Wei, S. D., & Lai, S. H. (2008). Fast template matching based on normalized cross correlation with adaptive multilevel winner update. *Image Processing, IEEE Transactions on* 17, 2227-2235.

A Semi-Automated System for Measuring the Effects of Technology Interventions in Older Adult Populations

Gbolabo Ogunmakin[1], Patricio Vela[1], Claudia Rebola[2]

[1]Georgia Institute of Technology, United States of America
[2]Rhode Island School of Design, United States of America

Introduction

Accurate behavioral analysis of people and their interactions is a time and labor intensive process requiring the efforts of trained operators to perform both observation and coding of a given social environment and its activities [4]. Video recording technology and annotation software can improve the accuracy by removing the real-time immediacy of the observation and coding processes. However, manual annotation through software does not fully resolve the time costs associated to coding behavior. This paper presents a semi-automated coding system that first performs coding through automated video analysis, then provides a software graphical user interface for manually correcting the automatically coded data. Since the automated system captures the majority of output required, the observer need only process a small fraction of the video. The net effect is to enable higher throughput behavioral analysis. The benefit of this semi-automated tool is that the final output is reliable, since it was vetted by a trained coder, and it enables the coding of long duration video sequences (days) within a few hours. Until automated video analysis can be made reliable, semi-automated coding tools may prove to be highly effective for performing long-term behavioral analysis.

Motivation

Development of the system was motivated by the desire to assess the long-term effects of social interventions in retirement communities. Loneliness has been found to be a cause of great discomfort amongst older adults in a retirement community [5]. Social interaction and social support impact quality of life and health, with studies showing that lack of either leads to higher mortality rates amongst lonely older adults [2,3]. Even though these communities try to provide social exposure, depression and isolation are still present due to residents not making use of the public spaces for socialization [6,8]. Caregivers are concerned with the effectiveness of the interventions they design to increase socialization. Measuring social interaction would provide helpful feedback on the effectiveness.

Evaluating the utilization and socialization in public spaces requires observation of the space and the interactions occurring within it. Researchers have used sensors such as RFID tags, magnetic head trackers, accelerometers, body sensors, and video cameras to determine events happening in a region of interest [5]. All of these sensors, except the video camera, can be intrusive since they usually need to be worn by the population of interest. The potential cost of retrofitting some of these environments could also be expensive. Video cameras are less intrusive and less expensive to install. The main drawback of using video is that it is time consuming to manually review and analyze the data collected. Automating the process is a feasible solution, but even with state of the art algorithms, there are still errors. The proposed tool automates the coding process, then provides an interface for manual post-processing by the coder to correct any mistakes made by the algorithms.

Automated Video Analysis Surveillance System

Currently, the video surveillance software processes the video from a single camera configured to view a scene where interactions could occur. Information about the layout of the scene and the camera's positioning within it is needed, which is then used to configure the parameters of the surveillance system. The surveillance system has several interconnected modules that progress from detection to event processing for generating the output coding. The first step in analyzing the video is modeling the background so that targets can be detected using

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

foreground detection. Once a target has been detected, its appearance model is learned for re-identification and tracking. The tracking result for each frame is used to perform foreground segmentation for updating the

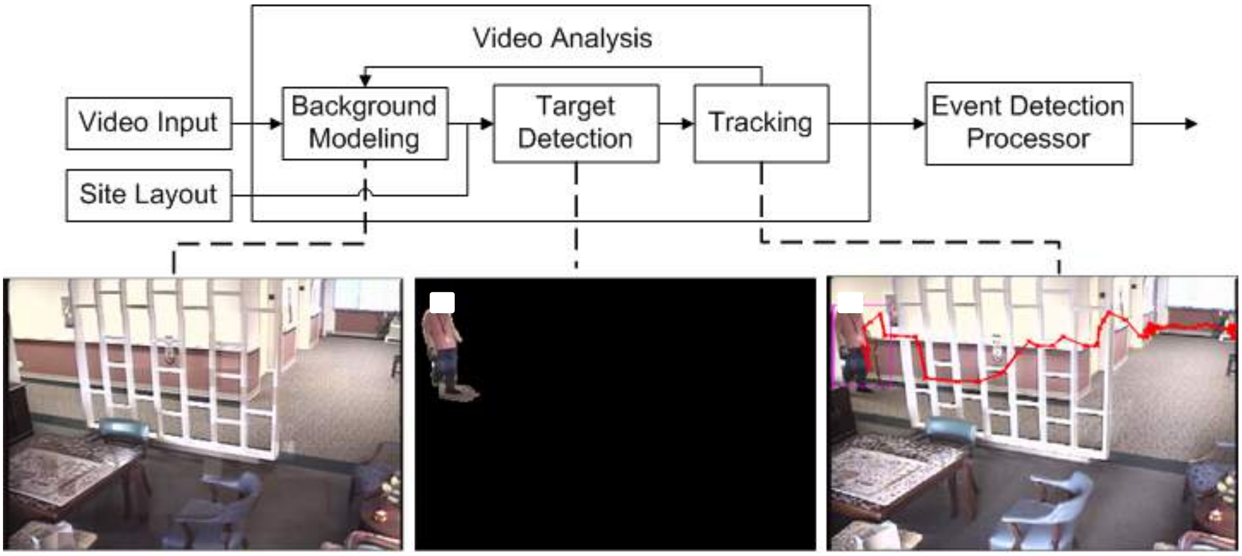


Figure 9. Process flow for the automatic surveillance system (face has been intentionally obscured).

background model and detecting new targets. Once the whole video sequence has been tracked, the results are passed to an event detection processor that calculates statistics such as the number of targets that entered the scene and the number of targets that interacted. An interaction is said to have occurred when two or more targets are stopped in close proxmity of each other for a given period of time.

The proposed system, as seen in Figure 1, uses a Gaussian mixture model (GMM) for foreground detection [7]. The GMM stores multi-modal representations of the background by modeling each pixel location using a mixture of *m* Gaussian distributions. A background estimation technique is utilized at three different intervals in the first video of the day being processed to create the initial *m* distributions for the background model. The probability of observing a pixel value, *x*, belonging to the background at time *t* is given by: $P(x_t) = \sum \omega_{i,t}\eta\left(x_t; \mu_{i,t}, \Sigma_{i,t}\right)$, where *m* is the number of distributions, $\omega_{i,t}$ is the weight of the i-th Gaussian of the mixture at time t, $\mu_{i,t}$ and $\Sigma_{i,t}$ are the mean and covariance of the i-th Gaussian at time t and η is the Gaussian probability density function. Foreground detection is performed by thresholding the probabilities of the pixels of the new frame belonging to the background. As the videos in each day are processed, the initial background model is updated to include gradual changes. Regions classified as foreground by the detector and the tracker are not included in the update.

Once a target is detected, it is tracked using a kernel covariance tracker [1]. The target is represented by a joint color-spatial feature vector utilizing the pixels in the detected region. The target's model is learned by mapping the feature vector, $u_i = [I(x_i), x_i]^T$, where $I(x_i)$ is the color data at location x_i, into a higher dimensional feature space using the Gaussian kernel, $\boldsymbol{k}(u_i, u_j) = \exp\left(-\frac{1}{2}(u_i - u_j)^T\Sigma^{-1}(u_i - u_j)\right)$. The eigenvectors, $\alpha^k = [\alpha_1^k, \dots, \alpha_N^k]$, and eigenvalues, λ^k, of the kernel matrix are computed. All the mapped points are then projected onto the normalized eigenvectors $f^k(u_i) = \sum_{j=1}^{N} \frac{\alpha_i^k}{\sqrt{\lambda^k}} \boldsymbol{k}(u_i, u_j)$. The target's new location is determined by using gradient descent to locally optimize the region similarity score $SC(\mathcal{R}) = \sum_{i=1}^{n}\sum_{k=1}^{ne}(f^k(u_i))^2$ (*n* is the number of feature vectors in the target's template and *ne* is the number of eigenvectors retained). The detected targets are tracked in this manner for every frame in the video and the resulting trajectories are passed to the event detection processor.

The custom programmed event detection processor takes the video analysis results and uses it to determine the number of targets that entered the scene and the interactions. The output of the event detection process is what an observer would normally create through annotations, then process to arrive at the outcome statistics. Given the layout, it determines the number of targets that were in close proximity to the intervention area. It also uses the targets' proximity to each other to determine whether an interaction is taking place. The results are then passed to a correctional graphical user interface that allows the user to view the results and make the corrections necessary. The hypothesis is that the amount of time required to process the data using the processed system is significantly less than manually reviewing and analyzing the data.

Graphical User Interface for Manual Correction

Given the fact that current state of the art visual processing and surveillance algorithms cannot achieve 100% accuracy, the ability to correct the resulting output is essential so a correctional GUI, as seen in Figure 2, was designed to review and correct the automated system's results. The user is able to enter the base path containing the folders for each day processed. A list of folders in the base path is loaded under the directories listbox, and whenever a folder is clicked, the videos for each hour of the day in the folder are displayed under the files listbox. Once a file is selected, it can be loaded using the "Load Video" button. The current file loaded is displayed right next the "Current File" textbox. The user then enters the name of the file where the automated results are stored and loads it using the "Load Results" button to get started.

Once the results are loaded, the "Targets List" listbox is populated with the targets detected throughout the day and the trajectories for all the targets are plotted. The time is also marked in order to know how long it takes to correct each day's results. When a target is selected, the video in which it was first detected is loaded, the current file is updated so that the user knows which file is currently being viewed, the slider is updated to the entrance point so that the user can hit the "Play" button to see how the target traverses the scene, its trajectory is plotted, and its entrance and exit frames into the scene and region of interest, along with its interactions list are displayed. Other targets that potentially match the selected targets are also displayed in the potential matches listbox. This allows the user to merge disjointed tracks of the same target.

The "Merge Tracks" button allows the user to merge two trajectory segments together, the "Split Tracks" button allows the user to remove an inconsistent trajectory segment from the target's history. The "Delete Track(s)" button allows the user to delete a track in case of a false positive. The "Interacting" button allows the user to declare an interaction between different targets that was not identified by the automated system. The "Interactions List" listbox shows the targets the selected target interacted with, and it allows the user to delete an erroneous interaction. The updated results are saved when the user selects a different day's folder, or when the user clicks the "Save" or "Quit" button.

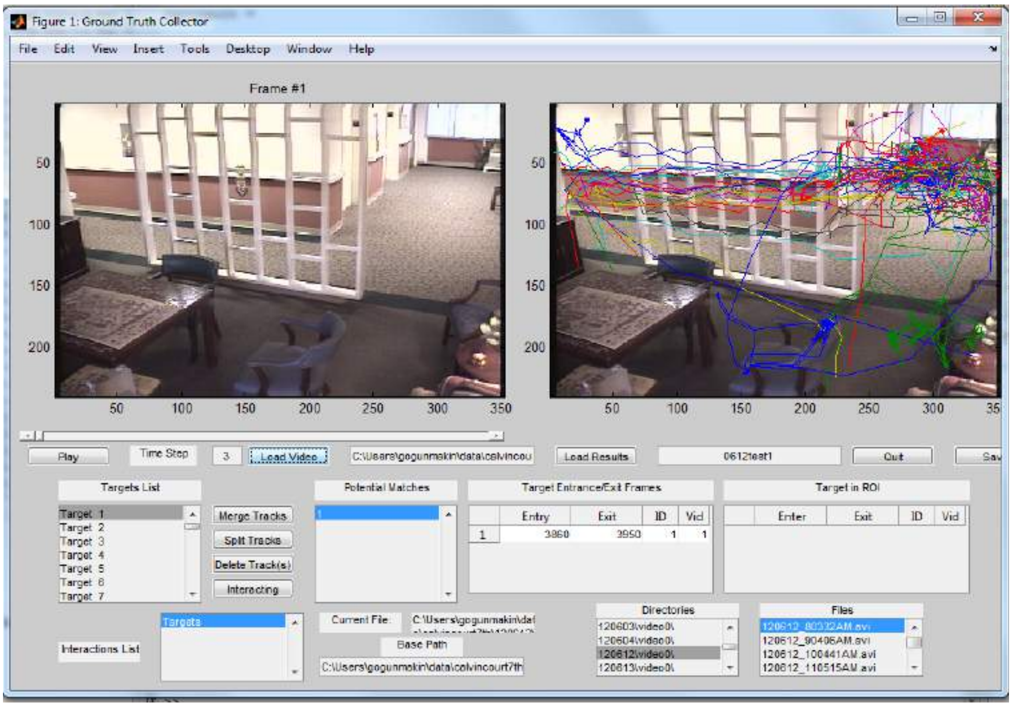


Figure 2. Graphical User Interface for Manual Correction.

Experimental Evaluation of the System

The motivation for the development of the proposed tool was to assess the long-term effects of an intervention in retirement communities. To determine whether the intervention is having the desired effect, behavior patterns with and without the intervention were monitored. To conduct an initial study on the impact of the intervention and the semi-automated coding tool, an iPad tower showing a slideshow of 180 images was installed at a local retirement community for adults aged 62 and older. The community is an 11-story building with all floors, similarly arranged, containing a common shared area near the elevators and long hallways for accessing individual apartments. Floors 2, 3, and 7 were monitored. A data collection script records the video and stores each hour in a different file in a folder named after the day it is recorded. A custom GUI is used to collect the ground truth for the number of people who entered the scene, the region of interest, and interacted [4].

The automated video analysis system is first utilized via a script containing the days and videos to process, and the layout information of the floor. A new background model is initialized at the beginning of each day, and updated as the rest of the day is processed. The targets' state at every frame is stored for use by the event detection processor. When all the videos in a particular day are processed, all the models are cleared, and the system processes the next day. After processing all the days in the script, the results are saved for the user to load and correct using the correction GUI.

The automated system detected and tracked 2523 target entries across the whole experiment, which is a 11.8% more than the 2258 entries recorded in the ground truth. The correction tool was used to correct the errors generated by the automated system to generate results with 100% accuracy. The automated system usually generated more false positives as a result of sudden illumination changes. Although the targets have their designated entrance regions, there are windows and light switches located in the common area. Whenever the illumination changes because of sudden cloud movement or someone turning on/off the light switch, the system sometimes erroneously detects and starts tracking a foreground object in the entrance region. This kind of error was corrected using the delete tracks button in the correction GUI. There were also some errors regarding when the targets left the scene. Sometimes, the system is not able to determine that the target has left the scene since it relies on the detection algorithm which sometimes produces false positives to due to the sudden illumination change not being incorporated into the model. This in turns creates false positives when it comes to detecting the number of interactions taking place. To fix this error, the “split tracks” button was used to split a target’s track at the moment it leaves the scene, the excess trajectory in the split track is then deleted.

Tables 1-3 shows the statistics of the three floors monitored. The total number of interactions for the first week is 16.4% of the total traffic and the total number of interactions for the second week is 10.9% of the total traffic. It should be noted that there was a medical emergency on Tuesday for the baseline condition of floor 3 which increased the number of total traffic and interactions. The total number of interactions for the baseline week of floor 3, excluding the anomaly, was 13.15% of the total traffic and the total number of interactions for the technology intervention week was 9.52% of the total traffic. The total number of interactions for the baseline week of floor 7 was 11.57% of the total traffic and the total number of interactions for the technology intervention week was 19.8% of the total traffic.

Table 4 shows the time it took to generate the event statistics using the custom GUI, Noldus Observer XT, and the Semi-Automated tool. The custom GUI time was computed by taking an average of the time it took to manually generate the ground truth results in Tables 1-3. The Observer time was computed by taking an average of the time it took to process 8 hours of the video using the Observer XT program. The semi-automated tool time shows the average time it takes to process the video automatically, and the time it takes for the user corrections. As hypothesized, the time required for the user to generate statistics using the semi-automated tool is significantly less than manual processing.

Table 1. Statistics of Baseline Floor (2nd).

	Total Daily Traffic				Total Daily Interactions			
	Week 1		Week 2		Week 1		Week 2	
	GT	Est.	GT	Est.	GT	Est.	GT	Est.
M	88	99	39	50	14	16	2	4
T	93	101	62	75	18	20	6	7
W	95	107	72	88	23	27	11	14
Th	52	67	74	87	4	7	4	6
F	44	51	74	83	2	4	12	14
	372	425	321	383	61	74	35	45

Table 2. Statistics of the Floor w/out Activity (3rd).

	Total Daily Traffic				Total Daily Interactions			
	Baseline		Technology		Baseline		Technology	
	GT	Est.	GT	Est.	GT	Est.	GT	Est.
M	47	55	44	50	13	15	0	2
T	100	117	63	66	29	34	10	12
W	74	79	61	66	6	8	2	5
Th	90	107	55	57	8	12	2	3
F	40	44	71	74	6	7	14	15
	351	402	294	313	62	76	28	37

Table 3. Statistics of Floor with Activity (7th)

	Total Daily Traffic				Total Daily Interactions			
	Baseline		Technology		Baseline		Technology	
	GT	Est.	GT	Est.	GT	Est.	GT	Est.
M	80	92	114	121	9	12	24	28
T	101	112	120	125	8	10	37	40
W	92	104	103	112	13	16	17	21
Th	62	67	80	88	8	10	8	11
F	80	84	88	95	10	11	14	18
	415	459	505	541	48	59	100	118

Table 4. Average time to code 60 minutes of video using acustm GUI, a general purpose coding software package, and the semi-automated tool.

Custom GUI	Noldus Observer XT	Semi-Automated Tool	
		User Corrections	Surveillance Program
22.62 minutes	73.8 minutes	5.8 minutes	65.7 minutes

Conclusion

This paper presented a tool for semi-automated coding of social interactions from recorded video. With regards to the desired output, the automated system was 88.26% correct in detecting and tracking targets thereby effectively requiring annotation of a significantly smaller portion of the overall video. With the system, a coder can process 4x more video than with a custom interface, and 13x more video versus with a general purpose software package, given the same amount of labor input. In particular, one 8-hour period of time can be coded in just over 45 minutes. Future work will be to improve the performance of the automated video analysis system so that its processing time is significantly reduced and its accuracy is increased, as well as to evaluate user interface modifications that could reduce the time spent during the correction step.

References

1. Arif, O., & Vela, P. A. (2009). Robust Density Comparison for Visual Tracking. In BMVC (pp. 1-10).
2. Elosua, P. (2011). Subjective values of quality of life dimensions in elderly people. A SEM preference model approach. Social indicators research, 104(3), 427-437.
3. Golden, J., Conroy, R. M., Bruce, I., Denihan, A., Greene, E., Kirby, M., & Lawlor, B. A. (2009). Loneliness, social support networks, mood and wellbeing in community-dwelling elderly. International journal of geriatric psychiatry, 24(7), 694-700.
4. Pisrek, J. And Modzelweski, M. (2014). Argos Observation System. Computer Program Enhancing Observation of Classroom Proceedings in Primary School. Proceedings of Measuring Behavior. Eds: AJ Spink, LWS Loijens, M Woloszynowska-Frase, and LPJJ Noldus, pages 363-365.
5. Rebola, C. B., Ogunmakin, G., & Vela, P. A. (2013). Design and Technologies for Understanding Older Adults Social Interactions in Retirement Communities. International Journal of Social Robotics, 5(4), 575-591.
6. St. John, P. D., & Montgomery, P. R. (2009). Do depressive symptoms predict mortality in older people?. Aging & mental health, 13(5), 674-681.
7. Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. (Vol. 2). IEEE.
8. Zimmerman, S., Mitchell, C. M., Chen, C. K., Morgan, L. A., Gruberbaldini, A. L., Sloane, P. D., ... & Munn, J. (2007). An observation of assisted living environments: Space use and behavior. Journal of Gerontological Social Work, 49(3), 185-203.

Behavioural Analysis of Mobile Web Users

Rami Albatal*, Peter Briggs, Maurice Coyle, Sebastian Gavarini, Johannes Tomin and Barry Smyth

HeyStaks Technologies Ltd., Ireland.
***rami.albatal@heystaks.com**

Abstract

As smartphones become the predominant devices for accessing the web, understanding how individuals express their interests and interact with the web can have a great impact on several domains ranging from customer services to marketing and public policy. However, in order to better understand the web surfing behaviour and interests of mobile network subscribers, we need to look beyond the classic analytics that are based on location, internet usage and social networks. A more granular view of user behaviour and interests can be achieved by including more advanced analytics based on the content that the users are engaging with. In this paper we present a novel mobile web content analytics platform, HeyStaks, with the goal of filling the gap of granular content analysis for mobile user behavioural analytics.

Introduction

It is clear that smartphones are increasingly becoming our main content consumption devices when it comes to accessing the web. According to the latest forecast from Cisco [1]: “Mobile data traffic will grow three times faster than fixed traffic in the period 2014-2019, driven by more devices and users as well as faster networks. [...] The growth is driven by more mobile users, which are expected to grow from nearly 59 percent of the world population last year to 69% in 2019.” Google has lately announced that “more Google searches take place on mobile devices than on computers in 10 countries.”

With these facts in mind, the analysis of web content engagements on mobile devices is of great interest for many market and social players. For example, Mobile Network Operators (MNOs) are constantly seeking to optimise their services with better recommendations and customised plans for their users, advertisers are looking for more relevant and less aggressive targeting mechanisms, and governments and non-governmental organisations want to better understand people's interests, trends and social engagement at population level.

We argue that a far more granular and accurate level of user profiling can be achieved by adding the essential element of web content analysis. It is cumbersome to track users across websites using cookies, and doing so requires partnerships with many content publishers in order to achieve a reasonable level of coverage, But the service provider has a global view of the user's mobile web activity, and with the user’s consent they can build a rich, anonymous profile of the user’s personal preferences. This is where the HeyStaks platform comes in to bridge the gap and enable MNOs to build a richer user profile.

This paper outlines the HeyStaks platform, and some key results from an analysis of the Web usage patterns of mobile subscribers. The HeyStaks platform extracts and identifies behavioural patterns from this usage data to better enable marketers, advertisers, and other parties to create precise targeted campaigns for interested audiences.

The rest of this paper is organised as follows: we first start by presenting related state-of-the-art works, then we describe HeyStaks and it’s methodology. Finally, we discuss some insights that we generated from a population of users from one of our partner MNOs.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Related Work

While the idea of investigating web logs is not new, there are few academic or industry research studies that combine the analysis of mobile web usage patterns with the automated profiling of mobile subscriber interests through the classification of the web content they access.

Many studies in the state-of-the-art have demonstrated the effectiveness of analysing web content and/or the URLs of the websites that users engage with. Hofman & Sirer[2] analysed the Web histories (by categorising the web pages into 5 categories) of 250,000 anonymized individuals with user-level demographic information. They examined how the surfing behavior changes as individual spend more time online, how it depends on educational background, and how browsing histories can be used to infer user's attributes. In [3] a method for predicting the dwell time on Web pages was proposed based on features related to the content (words), the HTML tags and measurements on the page size, heights and width and others. In [4], it was possible to predict some of the demographic information of users using their URLs visits patterns. Authors in [5] applied a clustering technique on URLs visited by the users to capture the common interests of different types of users.

When looking at mobile user behavioural analytics, most of the research has focused on one or a combination of three dimensions: social networks, mobility, and Internet usage metrics. For example, the study in [6] focused on the analytics of the user behaviour change over different networks using network traffic data. In [7], sub-communities of similar users are identified based on mobility and network traffic pattern analysis. Anindya & Sang [8] measure User content generation and usage behaviors based on factors related to calling patterns and mobility. A data mining method (SMAP-Mine) was proposed in [9] to discover the sequential movement patterns associated with requested services (e.g. restaurants, theatre).

Few studies used the domain names of the websites visited by a mobile user to perform behavioural and interest analysis. The authors in [10], proposes a probabilistic model hat combines the user location and the user interest profile generated by applying Latent Dirichlet Allocation over a bag-of-websites (domain names) representation, the model was used to perform collective behavioural analysis for mobile usage prediction and service recommendation. In [11], a profiling and recommendation approach based on fuzzy clustering is applied on the URLs visited.

The studies, and many others, can be very beneficial in enhancing the quality of user targeting and recommendation systems, however none of them apply advanced analytics on the content of the requested web pages. This is where the HeyStaks platform contributes to this domain by carrying out a detailed content analysis on the visited web pages to provide a richer user profile.

In the next section we will provide a high-level overview of HeyStaks and its content analysis methodology.

HeyStaks Big Data Platform for Mobile Web Behavioural Analytics

Figure 1. shows a high-level overview of the HeyStaks platform for mobile Web behavioural analytics.

The platform constitutes of two parts:

- First part (left side of Figure 1.): installed in the MNO Network. This part is mainly responsible for building, managing and updating the subscribers profiles. URL logs containing timestamped URLs with user ID and Location are read and parsed by the *Update API*. The Profile Manager module takes charge of aggregating the information related to each subscriber, and it requests the topics related to each URL from the *Topic Manager* module. The user profiles are then saved in a database that can be accessed via the *Access API*. The profiles can be further analysed by the *Community Manager* module.
- Second part (right side of Figure 1.): this part is outside the MNO network and contains the core analysis capabilities of the HeyStaks platform. It receives requests from the *Topic Manager* module that ask for the topics related to new URLs that the system has not encountered before. In this case, the

communication is conducted via a secure connection. The cloud service will fetch the content of the URL and analyse it via Machine Learning models. The Machine Learning models map the content of a URL into one or several topics in a taxonomy.

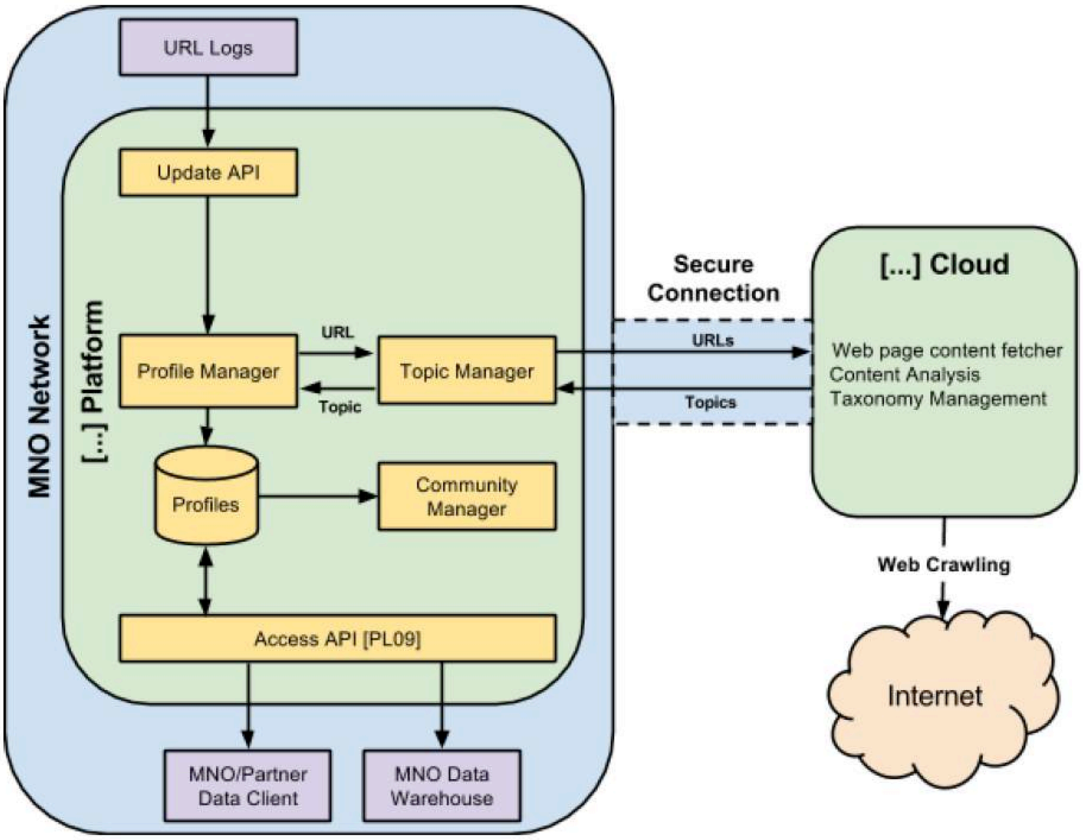


Figure 1. Overview of the HeyStaks platform for mobile Web behavioural analytics

Next we will provide some details about the modules that affect or contribute to the behavioural modelling of users.

HeyStaks Taxonomy Management

HeyStaks maintains a Taxonomy of categories that are used to classify the high-level interests of end users, and which is based on the IAB Quality Assurance Guidelines Taxonomy [12].

While the HeyStaks Taxonomy has been carefully designed to meet the needs of most use-cases for HeyStaks Profile data, some customers have specific requirements in this area. To accommodate these differing use-cases, HeyStaks supports alterations to the standard Taxonomy on a per-customer basis to cater for more localised interest topics. The Taxonomy may also be extended to allow for alternative use-cases for HeyStaks Profiles. For example, an eCommerce-focussed taxonomy could be constructed to meet the profiling requirements of an online store.

HeyStaks Web Content Analysis

The techniques used to determine the topics of the web pages include a structural analysis of the page content, link graph analysis of the pages, text relevance analysis, page term statistics, and intent analysis of search queries and page characteristics. The variety of content analysis techniques used ensure a high quality mapping between the web pages visited by the users and the topics in the HeyStaks taxonomy, and allow the system to perform well in a multilingual web environment.

First, the text processing step applies a standard stop-words list (available on http://www.ranks.nl/stopwords). Next, page features are extracted and mapped to an index using a hash function. This allows term frequencies to be calculated for the page features based on the mapped indices. This approach avoids the need to compute a global term-to-index map, which can be expensive for a large corpus. Along with Term Frequency/Inverted Document Frequency (TF-IDF) features [14], Machine Learning algorithms are used to analyse the textual content of the web pages. Specifically, Latent Dirichlet Allocation (LDA) [13] is employed for feature extraction. Textual features are then used as input for Naive Bayes and Logistic Regression multi-class classifiers. To train these classification algorithms, a ground-truth dataset is automatically constructed using common web search engines and open knowledge-bases such as WikiPedia. Relevant pages for each topic in the taxonomy are collected, then the pages are processed and features are extracted and used as input for the training algorithm. The generated classification models are then used to estimate which taxonomy topics are most likely to be related to a web page, then the pages are filtered by comparing them to reference pages on WikiPedia. The classification quality is measured using an "Accuracy@1" metric, and it is calculated over a standard 80%-20% split of the data into training and validation sets respectively. The system achieves an Accuracy@1 of 78% for pages with English content, and classifiers for several additional languages are currently being evaluated.

HeyStaks Community Manager

After analysing the web pages that a user visit, the results of the analysis are aggregated in the user profile and saved in *Profiles* database. The *Community Manager* module carries out offline processing on the HeyStaks *Profiles* database to identify communities of similar profiles within the MNO's subscriber population. Sets of similar users are computed using several clustering techniques, and these sets are then used to generate community-level profiles that aggregate the information in the underlying profiles, and also as input to the processing carried out by the Interest Inference Manager. In the next section we will present some of the insights that are produced by the HeyStaks *Community Manager* module.

Insights

Let's take an example of one day of web interaction data that belongs to a sample of 45K unique users from a mobile operator with approximately 3M subscribers. We will present 2 types of insights that we can generate from this sample.

A. Community Identification

In Table 1. below we see 3 examples of communities that were identified within the Web usage logs of a mobile operator. There are clear patterns of overlap in each community, indicating that the members are quite similar.

	Community 1	Community 2	Community 3
Primary interests	Business News	Clothing	Tourism
	Marketing	Women's Fashion	Accommodation
	Business Services	Footwear	Resorts
	Business Operations	Beauty	Vacation
Secondary interests	Business Logistics	Fashion Accessories	Social Science
	Wholesale	Health & Fitness Products	Computer programming
	Education	Men's Fashion	Travel Agencies
	Business Training	Cosmetics	Sport Events
	Business Associations	Fashion Designers & Luxury	Multimedia
	Newspapers & Mags	Skin Care	Computer Science
	Commerce	Shopping collectibles	Video Games
	Movies	Tourism	Consumer Electronics
	Tourism	Hair Styling	Religion

Table 1. Mobile Users Communities detected based on content analysis of Web usage

This fully automated processing of the user profiles reveals interesting communities that can be understood, described and can even be used to infer demographic information about the subscribers. It is clear from looking at the short profile excerpts in Table 1. that certain trends emerge - for example when we look at the 'Primary interests' of the communities, it looks like 'Community 1' consists of people who are interested in Business, 'Community 2' contains people who are interested in fashion and beauty, and 'Community 3' contains people who are interested in travelling.

B. Web Usage Behaviour

Using the Community Manager module, we can explore the web usage behaviour of the totality of the user population or we can decompose it into the usage behaviour of any discovered community. The graph in Figure 2. presents the global web usage behaviour in terms of number of engagements distributed over the 24 hours of a typical day.

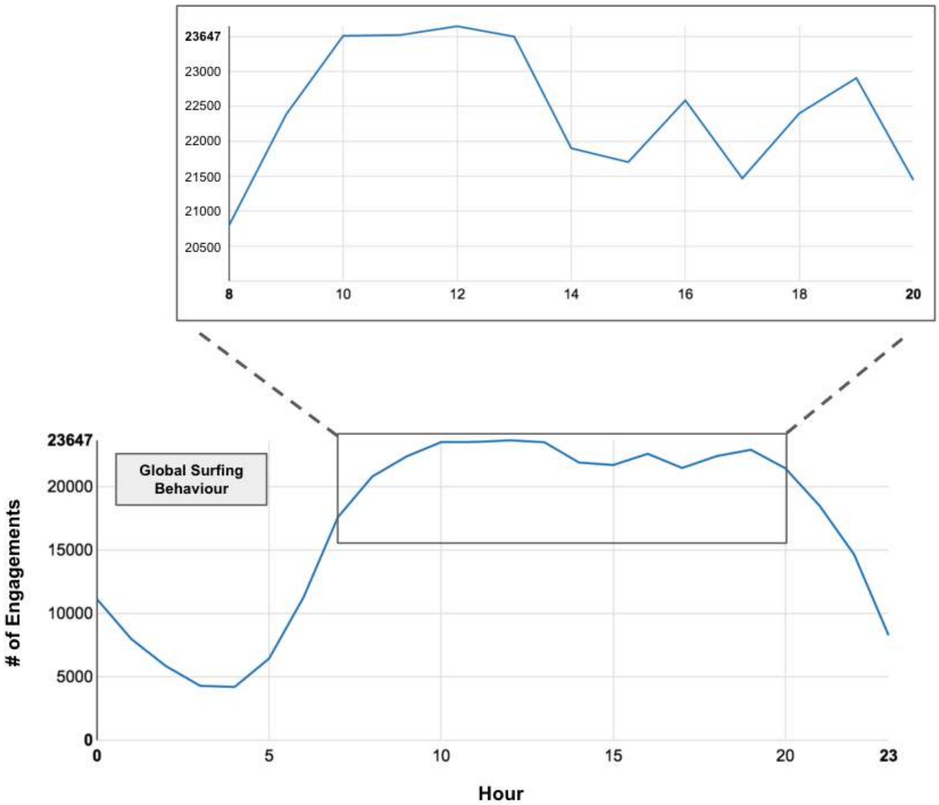


Figure 2. Global Web Usage Behaviour of a Population of MNO subscriber

What we notice is that across the entire user population there is a consistent level of usage during peak hours (between 8am and 8pm), and that there are a couple of particularly popular hours (12pm, 4pm and 7pm) that appear as peaks compared to their adjacent hours.

When looking at the communities in the Web Usage Behaviour Graphs in Figure 3, we can easily notice the differences. This information can play a crucial role in identifying the optimal time of day to start an advertising campaign, for example. The advertiser can now harmonise the timing of his targeted ads with the interest graphs of his target audience so that people see relevant ads when they are most likely to be useful and least likely to be an annoyance.

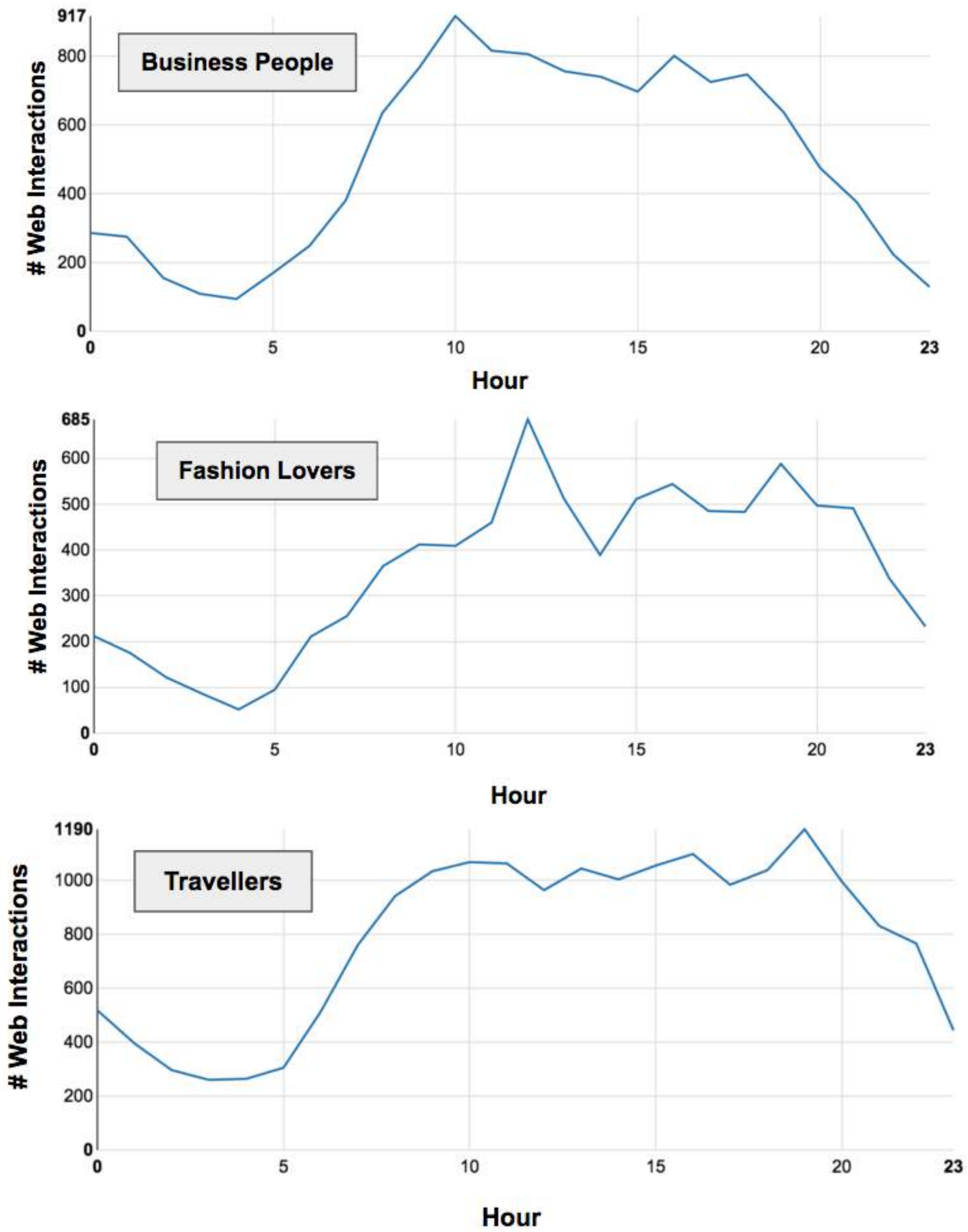


Figure 3. Surfing Behaviour Graphs of different MNO subscribers communities

It is interesting to see how different communities behave in different ways throughout the day. 'Business People' tend to read the business news while having their morning coffee or when they arrive at work. Many of our 'Fashion Lovers' check the Fashion websites and content during lunch, hoping to have some time in the evening to do some shopping. The 'Fashion Lovers' is a group that exhibit an intent to buy clothing, so they may respond positively to ads or offers for clothing deals around lunch time. After getting home from work, the 'Travellers' like to plan their trips and check promising holiday destinations in the evening, so this is the right time to target them with holiday ads and tickets promotions.

Ethical aspect of the study

The web usage data that was analysed in the course of this study was pseudonymized, which involved pre-processing the data to encrypt the user identifiers with a hashing function. The study was conducted by calculating aggregate statistics over the usage patterns and interests of the entire user population, and did not involve the analysis of individual users' data.

Conclusion

In this paper we presented some modules of the HeyStaks Mobile Data Analytics Platform. We have described the main modules that are used to analyse the web pages, to construct user profiles, and to analyse the collective behaviour of users communities. The interesting behavioural insights generated by the HeyStaks platform can allow MNOs and other interested parties to know who their subscribers are, how they engage with the web and when they are likely to be interested in specific topics.

References

1. "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper". Cisco Systems. February 3, 2015. Link. Published 3 February 2015. Accessed 27 January 2016.
2. Goel, S.; Hofman, J. M. & Sirer, M. I. (2012). Who Does What on the Web: A Large-Scale Study of Browsing Behavior. In proceedings of the 6th International AAAI Conference on Weblogs and Social Media (Dublin, June 2012), 1-8.
3. Liu, C., White, R.W., Dumais, S. (2010). Understanding web browsing behaviors through Weibull analysis of dwell time. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10 (New York, NY, USA, 2010), 379-386.
4. Adar, E., Adamic, L. A., Chen, F. R. (2007). User profile classification by web usage analysis. Google Patents. http://www.google.com/patents/US7162522.
5. Mobasher, B., Cooley, R., Srivastava, J. (1999). Creating adaptive web sites through usage-based clustering of urls. In Proceedings of the 1999 Workshop on Knowledge and Data Engineering Exchange - KDEX 1999 (Chicago, Illinois, USA, 1999), 19-25.
6. Tsompanidis, I., Zahran, A.H., Sreenan C. (2014). Mobile Network Traffic: a User Behaviour Model. In Proceedings of 7th IFIP Wireless and Mobile Networking Conference (Portugal, May 2014), 1-8.
7. Tang, D., Baker, M. (2000). Analysis of a local-area wireless network. In Proceedings of the 6th annual international conference on Mobile computing and networking (NY, USA, 2000), 1-10.
8. Anindya, G., Sang P.H. (2011). An Empirical Analysis of User Content Generation and Usage Behavior on the Mobile Internet. Management Science, 57(9), 2011, 1671-1691.
9. Tseng, V.S., Lin, K.W. (2006). Efficient mining and prediction of user behavior patterns in mobile web systems. Inf. Softw. Technol. 48, 6 (June 2006), 357-369.
10. Yua, B., Xu, B., Wu, C., Ma, Y. (2014). Mobile Web User Behavior Modeling. In Proceedings of the 15th International Conference on Web Information Systems Engineering - Wise 2014 (Thessaloniki, Greece, 2014), 388-397.
11. Phatak, D., Mulvaney, R. (2002). Clustering for personalized mobile web usage. In Proceedings of the 2002 IEEE International Conference on Fuzzy Systems - FUZZ-IEEE 2002 (Budapest, Hungary, 2002), 705-710.
12. IAB Quality Assurance Guidelines (QAG) Taxonomy. http://www.iab.com/guidelines/iab-quality-assurance-guidelines-qag-taxonomy/. Accessed 27 January 2016.
13. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (March 2003), 993-1022.
14. tf-idf. https://en.wikipedia.org/wiki/Tf-idf. Accessed 27 January 2016.

Expanding the Study of Internet Gambling Behavior: Patterns and Trends within the Icelandic Lottery and Sports Betting Platform

Jonsson, G.K. [1], Milkman, H. [2], Gray, H.M. [3], PaPlante, D.A. [3] and Shaffer, H.J. [3]

[1]Human Behavior Laboratory, University of Iceland, Iceland gjonsson@hi.is
[2] Metropolitan State University of Denver & [3] Harvard Medical School

Abstract

As rates of Internet gambling participation increase worldwide, so too does the need to understand how people engage in this form of gambling. This study represents the first examination of actual Internet gambling records within Iceland, a Nordic country with an active Internet lottery market that imposes strict regulations on gambling operator licenses. We summarized electronic betting records of a cohort of subscribers to the Internet betting service provider Íslensk Getspá. In addition we searched for temporal patterns in individual gambling records. We observed that the typical subscriber bet approximately 3 days per month and made fewer than two bets per gambling day, each worth approximately the equivalent of $4 U.S. Subscribers lost the bulk (96 %) of the amount they wagered, for a total loss of approximately $40 U.S. across the 2-year window of observation. Although these observations do not support the view of Internet gambling as an activity that is inherently risky for the typical subscriber, we did observe discontinuity across the distributions of gambling behavior, with the top 1 % of subscribers making more than three bets per day. Certain betting patterns were exclusively detected with frequent gamblers and different pattern types between gender and age groups.

Introduction

Several recent studies gathered convenience samples of Internet gamblers and collected self-reports about their behavior. Yani-de-Soriano et al. [1] estimated that 60 % of their sample of UK college students who reported past-month Internet gambling met criteria for at-risk or pathological gambling, and the at-risk and pathological gambling groups reported spending 5 and 7 hours per day gambling online, respectively. These groups of gamblers also reported experiencing a host of academic, health, and relationship problems, which could presumably result from spending so much time gambling online. Rather than using a self–selected sample LaBrie et al. [2] avoided some of the limitations of previous studies by (1) including 100 % of subscribers from a cohort who registered with the www.bwin.com (formerly Bet and Win) online gambling site and (2) relying on electronic betting records rather than self-report to summarize betting activity. The median values for fixed odds sports betting revealed that the typical subscriber placed only 2.5 bets per betting day, each worth €4, approximately 7 days per month. Subscribers who played poker participated in a median of one poker session every 3 days [3], and those who played casino-style games bet on casino games about once every 2 weeks [4]. However, each of these studies also revealed highly skewed distributions for indices of betting activity, suggesting that a minority of subscribers gambled in an intense fashion and might be experiencing gambling problems.

Auer and Griffiths [5] examined records from an operator that offers web-based lottery, casino, and poker to Austrian citizens with a focus on subscribers who set voluntary limits on the time and money they spend on the site. Dragicevic et al. [6] examined the intensity with which subscribers to three Internet gambling operators licensed in Malta engaged in casino-style games (e.g., roulette, slot games, table games. The largest cluster of subscribers, representing 79 % of the sample, averaged 4 days of activity and 3.7 bets per day during their first month of play. Gainsbury et al. [7] analyzed records of multi-channel (i.e., Internet, telephone, on-course) horse and sports betting from an Australian operator. Across products, consistent with findings reported by LaBrie et al. [2]: a) The number of bets was highly skewed in the direction of fewer bets; b) Most subscribers bet only a few times during a 9-year window of observation; and c) a small minority made many bets.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Somewhat different median value trends compared to LaBrie et al. [2] were observed. The Australian cohort made larger bets but bet less frequently (median frequency value was less than one day per month, compared to ~7 days per month); median bet value was equivalent to €6 (compared to €4). These differences might reflect differences in product types (i.e., sports betting versus horse betting) or might be due to other unknown variables, such as culture-specific gambling norms and operator-specific responsible gambling features.

The generalizability of previous findings regarding the intensity of typical Internet gambling must remain in question until researchers explore trends within more groups of players and more gambling operators

Aims of the study

The regulatory environment of gambling in Iceland, and the dominance of the lottery platform, distinguishes this market from those that have received the bulk of attention within the Internet gambling literature. In this study, we examine patterns of gambling behavior among a cohort of Icelandic residents who subscribed to Íslensk Getspá during January 2010.

We seek to answer the following specific research questions:

- How did this research cohort gamble online as measured by a number of previously-used indices of gambling activity (e.g., sum of active betting days, frequency of active betting days, bets per day)?
- Is there a definable sub-group of subscribers who gamble in a relatively intense fashion, and how intensely do these subscribers gamble?
- Are there significant variations in gambling patterns across the product types offered by Íslensk Getspá during the window of observation?
- Do males and females evidence different play patterns across these product types?

Materials and methods

This research cohort includes all 520 individuals who registered with Íslensk Getspá, the Icelandic National lottery, during January 2010, but it should not necessarily be considered representative of all Íslensk Getspá subscribers, or necessarily representative of all Iceland residents who gamble online. We prospectively tracked gambling behavior for this cohort from January 1, 2010 through December 31, 2011. The research software Theme [8, 9] was used to detect temporal patterns in individual betting records.

Measures

The available demographic characteristics of the research sample included year of birth and gender. The gambling behavior measures used in this report are based on daily aggregates of participants' actual betting activity records (in Icelandic krona also reported in this study in US dollar amounts).

We obtained, for each product type and across all products, four indices of betting activity:

(1) the sum of active betting days;
(2) the sum of monies wagered;
(3) the sum of bets; and
(4) the sum of monies lost

Using these four variables as well as the first and last active betting day for each product, we calculated five additional variables for each product type and across all games:

(5) duration of betting activity (number of days between the first and last active betting days, inclusive).

(6) frequency of betting activity (total number of active betting days divided by by the duration of play).

(7) bets per day (the total number of bets divided by the number of active betting days.

(8) krona per bet (the sum of monies wagered divided by the number of bets.

(9) percent lost (total amount of monies lost divided by the sum of monies wagered).

We also identified whether each subscriber engaged in each of the eight product types. Engagement in a particular game was defined as having at least one active gambling day during the period of observation. Thus the total number of games played for each subscriber was calculated across the 2-year window of observation.

Results

On average, participants in the full sample (n = 520) were 40.46 years old (SD = 14.23) when they subscribed to Íslensk Getspá during January 2010. The range of reported age was 5–84 years. Illogical values for age might have resulted from the appropriation of family members' national registry number during subscription. The sample was mostly male (68.5 %), with 31.1 % identified as female and 0.4 % having no gender data. On average, participants evidenced 22.42 active betting days (Median = 7.00) from a total possible 730 days (i.e., January 1, 2010 until December 31, 2011). During this window of observation, the average total sum of wagers was $241 USD (Median = 56 USD). On average, participants made a total of 46.51 bets (Median = 10.50); and lost a total of $171 USD (Median = 40 USD). The mean duration of participants' was 316.28 days (Median = 148.50). The average frequency of gambling activity was 32 % (Median = 11 %). Participants made, on average, 1.87 bets per active betting day (Median = 1.67) and their average bets were worth $7.28, each (Median = $4 USD). Finally, they lost an average of 80 % of the amount wagered (Median = 96 %).

Note that many of the variables, particularly the sum of monies wagered, sum of bets, bets per betting day, and money per bet are highly skewed in the direction lower values. Gender differences and age trends were examined in the nine cross-product betting variables. Compared to women, men wagered more, made more total bets, lost more, made bigger bets, and lost less money per bet.

Two trends appeared in the relationship between age (in years) and betting activity, though both were modest. Age was negatively related to bets per day and positively related to percent lost. The typical subscriber engaged in 2.55 products during the window of observation. Men and women did not differ in the number of games played. Age was negatively correlated with the sum of games played; younger players tended to play more games.

We defined the most heavily involved betters as the top 1 % of the sample in terms of the total amount wagered and total number of bets. These two groups, which are somewhat overlapping, each consist of four subscribers. Examining median values, we observed substantially larger values among the most involved bettors compared to the remaining sample in terms of the five aggregate variables (i.e., total active betting days, total amount wagered, sum of bets, total amount lost, and duration of betting activity).

The two most involved betting groups also made more bets per day: the median values were 3.79 and 3.26 for the groups defined by large wager sums and large number of bets, respectively, and the median value for the remaining sample was less than two. Both intense betting groups lost a smaller proportion of their wagers than the remaining sample. However, the intense betting groups differed from each other: a) The group defined by large wager sums had a similar frequency of betting activity and made larger bets compared with the remaining sample; and b) The group defined by number of bets had a larger frequency of betting activity and made smaller bets than the remaining sample.

The most popular product in terms of number of active players was a lottery game with a weekly draw (i.e., ''VikingLotto´ 6/48''). Two hundred and eighteen players made at least one purchase on this game. The lottery games tended to be more popular than the sportsbetting games in terms of total number of active players. The least popular game was a type of live-action sportsbetting (n = 30). Expenditure per bet tended to be lower for lottery games and their add-ons than sportsbetting games. Frequency of betting activity ranged from 31 % of active betting days (VikingLotto´ 6/48) to 53 % of active betting days (Tippað ı´ beinni/LiveBetting). A greater percentage of women than men engaged in the four traditional lottery games (i.e., Lotto´ 5/40, VikingLotto´ 6/48, Joker Lotto´ 5/40, and Joker VikingLotto´ 6/48. By contrast, men were more likely than women to engage in the sports games, though this difference did not reach the adjusted level of significance ($p<.006$) for any sports game except English soccer betting. Engagement in each of the eight products was studied as related to subscribers' age: Those who engaged in Lengjan, the daily sportsbetting game, were younger (Mean age = 31.28 years); Those who engaged in the live-betting game (Tippað í beinni) were younger (Mean age = 31.28 years). The most involved subscribers (i.e., in terms of total amount of money wagered) played more games (Mean = 5.75) compared to their counterparts who wagered less (Mean = 2.51). The most heavily involved sub-group was more likely to engage in all products than the less involved subscribers, but this was especially pronounced for European soccer betting; 75 % of the top 1 % on money wagered played this game, compared to 9.4 % of the remaining sample. Similarly, those who made the most bets played roughly twice as many games as their counterparts and were especially more likely to engage in European soccer betting.

Certain betting temporal patterns were exclusively detected with frequent gamblers and different pattern types between gender and age groups.

Discussion

The primary goal of this work was to describe Internet gambling activities among a cohort of Internet lottery subscribers within a gambling environment that has not received empirical attention. Inspection of cross-product median values indicates that the typical subscriber bet approximately 3 days per month and made fewer than two bets per gambling day, each worth approximately $4 US. Subscribers lost the bulk (96 %) of their wagers, for a total loss of approximately $40 across the two-year window of observation. These observations do not support the view of Internet gambling as an activity that is inherently more risky than other forms of gambling.

The methodological shortcomings of other studies, particularly the use of self-selected convenience samples and self-reports rather than electronic records of gambling behavior, might have contributed to overestimates of the time and money the typical subscriber spends on Internet gambling. Nevertheless, some individuals who gamble online experience problems with their gambling. To the extent that this group is overrepresented in studies using self-selected samples, estimates of typical Internet gambling behavior will be biased—particularly if these players overestimate the intensity of their own gambling behavior.

Another goal of the current work was to collect descriptive information about Internet gambling that could be compared to work collected with other samples and within other betting environments. The body of evidence regarding actual Internet gambling includes some potentially interesting differences across cohorts; there is an emerging consensus that the majority of Internet subscribers gamble in a moderate fashion.

Broadening the scope of research into internet gambling within a Nordic country

Despite many restrictions, gambling in Iceland is quite popular; the most recent estimates suggest that 76 % of adults participated in gambling at least once during the past year. During Internet connections, the period 2002–2006, many more Icelandic residents began participating in Internet gambling, due, in large part, to an increase in the proportion of homes with high-speed internet connections. During 2011, 19 % of Icelandic survey respondents reported gambling online during the past year; 3 % reported gambling on foreign websites.

As in most countries, only a fraction of Icelandic citizens who engage in gambling report experiencing serious gambling-related consequences; during 2011, when asked about consequences experienced during the past year, approximately 0.8 % of the adult population reported meeting clinical criteria for a gambling disorder (i.e., Level 3 gambling disorder) and an additional 1.7 % evidenced subclinical symptoms (i.e., Level 2 gambling disorder)

The combined prevalence rate of 2.5 % for Level 2 and Level 3 gambling represents an increase from previous years (i.e., 1.6 % in 2005 and in 2007).

Compared to a recent pan-European estimate covering prevalence estimates gathered during the years 1997–2010, the 2011 Icelandic rate of Level 3 (more severe) gambling disorder is fairly representative of European countries (i.e., 0.8 % in Iceland compared to 0.6 % across Europe), although the 2011 Icelandic rate of Level 2 gambling disorder is somewhat higher (i.e., 1.7 % in Iceland compared to 1.0 % across Europe).

Currently, efforts to minimize gambling-related public harms center on legal restrictions (e.g., advertising restrictions, mandated use of funds for research and public health, few new licensing opportunities), screening for gambling disorder among patients attending treatment for substance use disorders, helplines (phone and web-based) operated by Gamblers Anonymous and Red Cross Iceland, and a responsible gambling website (http://www.abyrgspilun.is/) operated by the primary Icelandic lottery operators.

Research Implications and Future Directions

- Electronically recorded betting indices, such as those used in the current study, overcome some limitations of self-report but do not provide insight into players' thoughts and feelings.
- A limitation of this work, therefore, is the lack of information regarding how subscribers feel about their gambling and its influence on their lives.
- Future research might triangulate on a fuller description of problematic Internet gambling using a combination of betting records and self-report measures.
- This study is also limited because it describes the betting behavior of subscribers to only one Internet gambler operator; results should be considered in relation to their consistency (or lack thereof) with other descriptions of Internet gambling activity.
- This study describes how Icelandic residents gambled online during a unique period of their country's history, as they recovered from a major economic and political crisis that included a sharp rise in unemployment, and sharp drops in the value of the krona and in gross domestic product.
- Evidence is beginning to emerge that being economically disadvantaged might contribute to risk-taking in general and gambling in particular.
- Further research is needed to determine the extent to which the patterns we observed in this study were unique to this period in Iceland's economic and political history.
- In future work, we plan to develop a more comprehensive understanding of this Icelandic cohort by collecting information about the use of responsible gambling features and other signs of gambling problems.
- This information will allow us to examine how the behavior of gamblers who have potential gambling problems differ from their counterparts without such problems, potentially in the form of a predictive algorithm that will permit more accurate and effective prevention and treatment programs.
- We hope to use this data source to conduct research that informs the development of prevention and treatment initiatives that will limit social harms associated with Internet gambling.

Reference

1. Yani-de-Soriano, M., Javed, U., & Yousafzai, S. (2012). Can an industry be socially responsible if its products harm consumers? The case of online gambling. Journal of Business Ethics, 110(4), 481–497.
2. LaBrie, R. A., LaPlante, D. A., Nelson, S. E., Schumann, A., & Shaffer, H. J. (2007). Assessing the playing field: A prospective longitudinal study of internet sports gambling behavior. Journal of Gambling Studies, 23(3), 347–362.
3. LaPlante, D. A., Kleschinsky, J. H., LaBrie, R. A., Nelson, S. E., & Shaffer, H. J. (2009). Sitting at the virtual poker table: A prospective epidemiological study of actual Internet poker gambling behavior. Computers in Human Behavior, 25, 711–717. doi:10.1016/j.chb.2008.12.027.

4. LaBrie, R. A., Kaplan, S. A., LaPlante, D. A., Nelson, S. E., & Shaffer, H. J. (2008). Inside the virtual casino: A prospective longitudinal study of Internet casino gambling. European Journal of Public Health, 18(4), 410–416.
5. Auer, M., & Griffiths, M. D. (2012). Voluntary limit setting and player choice in most intense online gamblers: An empirical study of gambling behaviour. Journal of Gambling Studies.
6. Dragicevic, S., Tsogas, G., & Kudic, A. (2011). Analysis of casino online gambling data in relation to behavioural risk markers for high-risk gambling and player protection. International Gambling Studies, 11(3), 377–391.
7. Gainsbury, S., Sadeque, S., Mizerski, R., & Blaszczynski, A. (2012a). Wagering in Australia: A retrospective behavioural analysis of betting patterns based on player account data. Journal of Gambling Business and Economics, 6(2), 50–68.
8. Casarrubea M, Jonsson G.K, Faulisi F, Sorbera F, Di Giovanni G, Benigno A, Crescimanno G, Magnusson M.S. (2015). T-pattern analysis for the study of temporal structure of animal and human behavior: A comprehensive review. J Neurosci Methods. pii: S0165-0270(14)00349-5.
doi: 10.1016/j.jneumeth.2014.09.024.
9. Magnusson, M. S. (2000). Discovering Hidden Time Patterns in Behavior: T-Patterns and their Detection. Behavior Research Methods, Instruments and Computers, 32(1): pp. 93-110.

New Effective Leader-Behavior Insights through Combining Skin Conductance with Video-coded Field Behaviors

A.M.G.M. Hoogeboom and C.P.M. Wilderom

CMOB, University of Twente, Enschede, the Netherlands. a.m.g.m.hoogeboom@utwente.nl

Recent social research and theorizing is increasingly addressing the meaningful biological components of effective human behavior, thus to fully understand the effect of complex leader behavior, it is important to gain insight into the underlying physiological processes (e.g., [1] [2] [5] [6] [7]). Leadership studies have only recently begun to examine the link between leader behavior and psychobiologic, neuro measures (e.g., [8] [9]). Several scholars noted the importance of understanding the neurobiological underpinnings of human/leader behavior (e.g., [10]), calling for more knowledge about "the neurological processes behind a leader's behavior" [11]. However, studies that actively combine psychological data with measures of leader behavior and its relative effectiveness remain scarce.

Additionally, scholars emphasized that perceptual measures of leader behavior do not reflect actual displayed behavior (e.g., [12] [13] [14] [15]). To obtain more objective insight into a leaders' behavioral repertoire, a video-observation method is employed (e.g., [16] [17] [18]). Other scholars noted the importance of not only adopting a "range of methodological approaches," but also strong theory-building to advance the leadership field [19]. Combining both video-observed and systematically coded leader behavior with EDA measures results in an advanced understanding of the effects of different behaviors. The interaction between biological/physiological measures and theoretically-driven organization-behavioral research is especially critical for the advancement of both neuroscience and the field of management or business administration [1].

The aim of the present study was to examine how a biological approach, using skin conductance/EDA measures, can inform leadership research as well as demonstrate methodological procedures for combining observed leader behaviors and skin conductance measures (using "The Observer XT" from Noldus Information Technologies). In addition, we explored the relation between physiological EDA/arousal, various leader behaviors and leader effectiveness. In terms of leader behaviors, the study relied on transformational-transactional theory; the initiating structure-consideration paradigm; and research on contra-productive leader behavior. Drawing upon these behavioral theories, a codebook was developed which contains 18 mutually exclusive behaviors (see Table 1 for the list of behaviors and definitions).

Participants were 39 formal leaders in regular, randomly selected, staff meetings from one large Publicsector organization in the Netherlands. Staff meetings were chosen due to the direct interaction with followers, because leadership is especially visible during regular staff meetings, also termed talk-in-interaction, and because the hierarchical position of the leader is clear in this work context [20] [21] [22] [23]. A random, stratified procedure was followed to select the leaders in the sample. Participation in the study was on a voluntary basis. All the leaders in the sample could also decide not to wear the Empatica E4 bracelet while being video-taped. However, everyone who participated in this study complied. At least one hour of skin conductance (SC)/Electrodermal Activity (EDA) measures from the focal leader was obtained during a regular staff meeting with their followers. In addition, expert and follower leader effectiveness ratings were collected. Video-observation based microanalyses of effective leader behavior were related and compared with lower versus higher levels of so-called Skin Conductance Level and Response (SCL and SCR). We will offer the precise results of our empirical examinations in the conference presentation: for an initial set of descriptive findings, see Figure 1.

NB. Because skin conductance varies between subjects, due to ambient temperature, skin temperature, humidity, age, gender, ethnicity or time of day, and in order to enable comparison between subjects, ranges from the 5th to the 95th percentile (as thresholds) were calculated [24] [25] [26] [27]. In this study, the 5th percentile is called

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

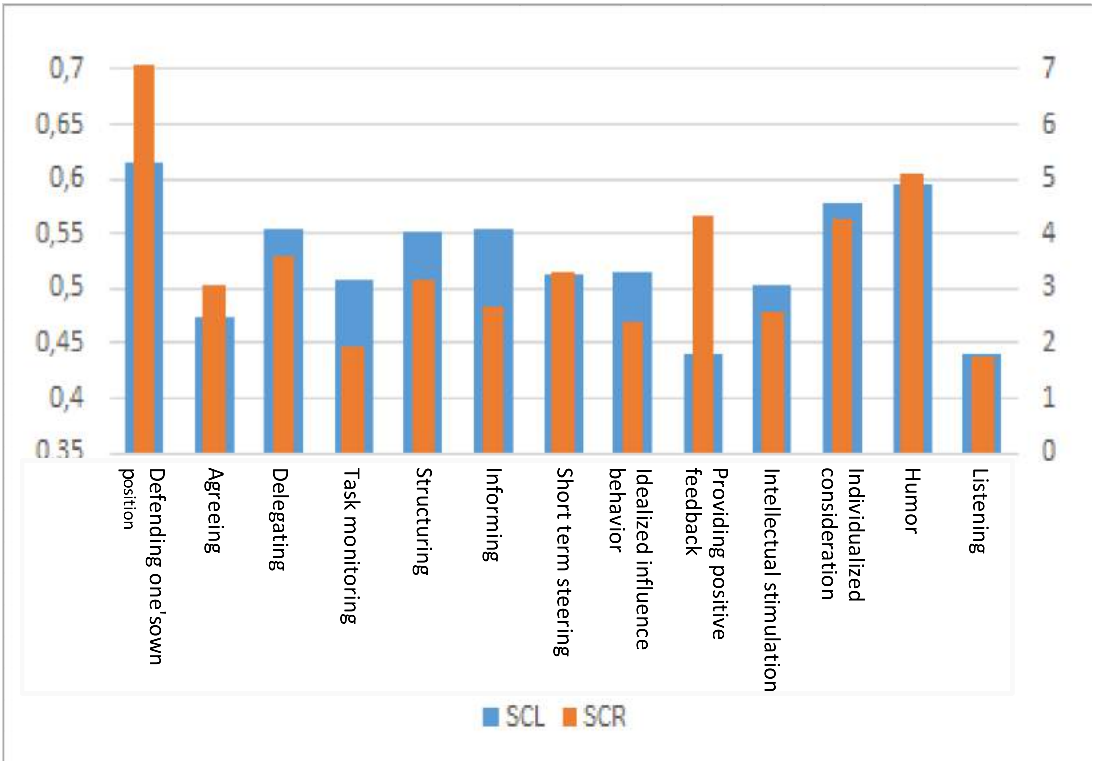
"0" and the 95th percentile is referred to as "1"; all other values are calculated within this range. In addition, skin conductance data is generally shown to increase after certain stimuli; changes in SCR/SCL are considered valid after 3 seconds [28] or even between 10 and 30 seconds [29]. SCL normally decreases at rest and rapidly increases after a new stimulus [28]. Hence, by combining the skin conductance data with the video-observed behaviors

(using a so-called "marker") SCL and SCR measures are used from 3 seconds after every start of a behavior and 1 second after a behavior ends. Some behaviors are not shown in Figure 1, because while some behaviors occur frequently, their duration is too short to validly use the SCR and SCL measures.

Table 1 Video-coded behaviors

Video-coded Behaviors	**Definition**
1. Showing disinterest	Not taking any action (when expected) Not listening actively (Counterproductive)
2. Defending one's own position	Emphasizing one's leadership position; Emphasizing self-importance (Counterproductive)
3. Interrupting	Shortly interrupting a team member when he or she is speaking (Counterproductive)
4. Disagreeing	Objecting or opposing other team members (Counterproductive)
5. Providing negative feedback	Criticizing the behavior of team members; Focusing on irregularities or mistakes (Transactional)
6. Task monitoring	Checking upon the task progress of team members' current situation; Referring to previously made agreements with team members; Making clear who is responsible (Transactional)
7. Correcting	Imposing someone to conform to norms or agreements (Transactional)
8. Delegating	Dividing tasks (without enforcing them); Determining the current direction (Initiating structure)
9. Informing	Giving factual information "The budget for this project is…" (Initiating structure)
10. Structuring	Structuring the meetings; Changing the topic; Shifting towards the next agenda point (Initiating structure)
11. Short term steering	Giving personal opinion about future plans (Task-oriented)
12. Idealized influence behavior	Specifying the importance of having a strong sense of purpose and talking about the collective mission of the organization (Transformational)
13. Intellectual stimulation	Positively stimulating the behavior of followers; Challenging professionally (Transformational)
14. Individualized consideration	Showing an interest in other's feelings, situation, or development; Showing empathy; Creating a friendly environment (Transformational)
15. Providing positive feedback	Giving complements or rewards (Relation-oriented)
16. Agreeing	Approving or supporting other team members (Relation-oriented)
17. Humor	Making friendly or funny jokes (Relation-oriented)
18. Giving personal information	Giving information on home situation or personal situation (non-work related) (Relationoriented)

Figure 1 Combination of video-observed leader behaviors with SCL and SCR measures



References

1. Lee N., Senior C., Butler M.J.R. (2012b). Leadership research and cognitive neuroscience: The state of the union. *The Leadership Quarterly* **23**, 213–218.
2. Marci, C.D., Ham, J., Moran, E., Orr, S.P. (2007). Physiological correlates of perceived therapist empathy and social-emotional process during psychotherapy. *The Journal of Nervous and Mental Disease* **195**, 103-111.
3. Spisak B.R., Homan, A.C., Grabo, A., Van Vugt, M. (2007). Facing the situation: Testing a biosocial contingency model of leadership in intergroup relations using masculine and feminine faces. *The Leadership Quarterly* **23**, 273-280.
4. Vrana, S.R., Gross, D. (2004). Reactions to facial expressions: Effects of social context and speech anxiety on responses to neutral, anger, and joy expressions. *Biological Psychology* **66**, 63–7.
5. Pastor, J.C., Mayo, M., Shamir, B. (2007). Adding fuel to fire: The impact of followers' arousal on ratings of charisma. *Journal of Applied Psychology* **92 (6)**, 1584–1596.
6. Heaphy, E.D., Dutton, J.E. (2008). Positive social interactions and the human body at work: Linking organizations and physiology. *Academy of Management Review* **33**, 137–162.
7. Hubert, W., De Jong-Meyer, R. (1990). Psychophysiological response patterns to positive and negative film stimuli. *Biological Psychology* **31**, 72-93.
8. Balthazard, P.A., Waldman, D.A., Thatcher, R.W., Hannah, S.T. (2012). Differentiating transformational and non-transformational leaders on the basis of neurological imaging. *The Leadership Quarterly* **23**, 244–258.
9. Molenberghs, P., Prochilo, G., Steffens, N.K., Zacher, H., Haslam, S.A. (2015). The neuroscience of inspirational leadership the importance of collective-oriented language and shared group membership. *Journal of Management*, 1–27.
10. Senior, C., Lee, N., Butler, M. (2011). Organizational cognitive neuroscience. *Organization Science* **22**, 804– 815.

11. Boyatzis, R.E., Passarelli, A.M., Koenig, K., Lowe, M., Mathew, B., Stoller, J.K., Phillips, M. (2012). Examination of the neural substrates activated in memories of experiences with resonant and dissonant leaders. *The Leadership Quarterly* **23**, 259-272.
12. Bono, J.E., Judge, T.A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology* **89**, 901–910.
13. Brown, D.J., Keeping, L.M. (2005). Elaborating the construct of transformational leadership: The role of affect. *The Leadership Quarterly* **16(2)**, 245-272.
14. Shondrick, S.J., Dinh, J.E., Lord, R.G. (2010). Developments in implicit leadership theory and cognitive science: Applications to improving measurement and understanding alternatives to hierarchical leadership. *The Leadership Quarterly* **21**, 959–978.
15. Dinh, J.E., Lord, R.G., Gardner, W.L., Meuser, J.D., Liden, R.C., Hu, J. (2014). Leadership theory and research in the new millennium: Current theoretical trends and changing perspectives. *The Leadership Quarterly* **25(1)**, 3662.
16. Alvarez, J.L., Svejenova, S. (2005). *Sharing Executive Power: Roles and Relationships at the Top*. Cambridge: Cambridge University Press.
17. Fairhurst, G.T., Uhl-Bien, M. (2012). Organizational discourse analysis (ODA): Examining leadership as a relational process. *The Leadership Quarterly* **23(6)**, 1043–1062.
18. Hoogeboom, A.M.G.M., Wilderom, C.P.M. (2015). Effective leader behaviors in regularly held staff meetings: Surveyed vs. Videotaped and Video-Coded Observations. In: J. A. Allen & N. Lehmann-Willenbrock & S. G. Rogelberg (Eds.), *The Cambridge Handbook of Meeting Science*. Cambridge University Press, 381-412.
19. Uhl-Bien, M., Riggio, R.E., Lowe, K.B., Carsten, M.K. (2014). Followership theory: A review and research agenda. *The Leadership Quarterly* **25(1)**, 83-104.
20. Fairhurst, G.T. (2007). *Discursive Leadership: In Conversation with Leadership Psychology*. Thousand Oaks: Sage.
21. Larsson, M., Lundholm, S. (2013). Talking work in a bank: A study of organizing properties of leadership in work interactions. *Human Relations* **66(8),** 1101-1129.
22. Allen, J.A., Yoerger, M.A., Lehmann-Willenbrock, N., Jones, J. (2015). Would you please stop that!? The relationship between counterproductive meeting behaviors, employee voice, and trust. *Journal of Management Development*, **34(10)**, 1272-1287.
23. Svennevig, J. (2008). Exploring leadership conversations. *Management Communication Quarterly* **21 (4)**, 529536.
24. Lykken, D.T. (1962). Range correction applied to heart rate and to GSR data. *Psychophysiology* **9**, 373-379.
25. Lykken, D.T., Venables, P.H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology* **8(5)**, 656-672.
26. Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in Skin Conductance. *Psychophysiology* **22(3)**, 292-299.
27. Westerink, J., Ouwerkerk, M., de Vries, G.J., De Waele, S. (2009). Emotion measurement platform for daily life situations. *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*.
28. Dawson, M.E., Schell, A.M., Filion, D.L. (2007). The electrodermal system. In Cacioppo, J.T., Tassinary, L.G., & Berntson, G. (Eds.), *Handbook of Psychophysiology*. Cambridge: Cambridge University Press, 157-181.
29. Boucsein, W. (2012). *Electrodermal Activity.* New York: Springer.

Altered Video Task: A Promising Alternative for Elicited/deferred Imitation Task in Young Children

Katarzyna Bobrowicz[1,2], Maciej Haman[2,] and Ryszard Bobrowicz[3]

1: Lund University, Lund, Sweden. katarzyna.bobrowicz@lucs.lu.se
2: University of Warsaw, Poland.
3: University of Copenhagen, Denmark.

Introduction

The method presented in this paper is invented to address the problem of episodic memory in participants with highly restricted verbal abilities, 15-month-olds in this case. Here, we refer to episodic memory defined as a mind/brain system with three main responsibilities: to encode, to store and to recall individual memories. Episodic memory system requires three additional capacities, which make it uniquely human: a sense of self, a sense of subjective time, and autonoetic awareness [26]. Such approach is currently predominant both in developmental psychology and in cognitive zoology research.

With a focus on developmental studies, we introduce a method, which aims to pair a measure of episodic memory and a measure of self-awareness. Episodic recall is measured via presentation of an original and a modified recording of a personal past event after a delay. The participant is expected to watch the unfamiliar video significantly longer than the familiar video, and so evince the differentiation between them. Alongside, the participant takes part in a mirror-mark task (a standard measure of self-recognition) and also in a real-time video task (a possible alternative for the mirror-mark task).

Measuring of the recall is based on "what"-"who"-"where" aspects of the past event. Three modified videos are generated from the original one, and the modifications refer to: 1. a toy ("what"), 2. an experimenter ("who") and 3. a setting ("where"). That is why this method also derives from cognitive zoology studies, where episodic memory is measured via behavioural signs of remembering "what", "where" and "when" happened [9]. The "who" aspect is a common addition in case of highly social animals [23][24][25].

The most typical method of measuring episodic recall in human children, even as young as 6-month-olds [2, p. 175] is elicited/deferred imitation task. However, it does not involve measuring of any of the "uniquely human" capacities and can be only applied to these organisms, which can readily imitate human experimenter's actions.

Further, we also elaborate on the above-mentioned issues. We also discuss the results and possible improvements of the method implementation, for we actually tested it with a group of 15-month-olds. The results were statistically significant for the "who" and "what" aspects, but remained insignificant for the aspect of "where".

State of knowledge

Since our method is proposed as an alternative for the elicited/deferred imitation task, we start with the concise description of the second. Imitation task is an implementation of a Piagetan idea, which points to imitation after delay as a reliable test of recall [2, p. 177]. This task includes three phases: 1. a participant may play with the props spontaneously; 2. the participant observes a model, who involves the props in a single action or a sequence of actions; 3. the participant performs spontaneously any actions with the props, either immediately (elicited imitation) or after a delay (deferred imitation) [2][3][8][17]. This method proved successful in children as young as 6-month-olds, when 24-hour delay was introduced [2]. For it has a strong theoretical background and is recognized as a valid method of measuring non-verbal recall, it is currently predominant in episodic memory research, which involves human children with highly restricted verbal abilities.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

What exactly is measured in the imitation task? It is based on the recall of “which prop”-“which action”-“in which order” was presented in the initial exposure. Although this recall is considered episodic, it does not necessarily fulfil three required capacities that accompany the episodic capacity in Tulving’s theory [26]. Measuring neither of a sense of self, nor of a sense of subjective time, nor of autonoetic consciousness is involved. Behavioural signs of the three pose a reasonable challenge, but we believe that a method, which could address both episodic capacity, and at least one of the others, may offer a better insight into their joint (or not joint) development. Aware that cognitive zoology research struggles with such change as well, we implement its approach towards episodic recall and accompany our procedure with the measures of autonoetic consciousness, which have already been described within developmental approach.

Since episodic memory is a probably uniquely human capacity [26, p. 5], the corresponding ability is named “episodic-like” within cognitive zoology research. Although it was initially understood as an integrated retrieval of “what”, “where” and “when” [9], these three aspects were soon accompanied by “who” [23][24][25]. „What” usually refers to food alteration, „where” to the change of food location, and „who” to the change of the experimenter present. „When” is particularly troublesome, since perception of time varies with subject’s biological constitution [18]. Nevertheless, it is usually operationalized as an order of events or actions. Each time food gratifications were excluded from the procedure, „where” and „when” have been recognized within animal cognition research as more problematic than „what” and „who”, since highly social individuals pay considerable attention to the agents (manipulated/self-moving props and individuals) in their environment. Food inclusion seems to be highly avoided in developmental studies with human participants, so we employed slightly different solutions for „where” and „when” aspects than the above-mentioned. In case of „where”, we referred to Gaffan’s [11] scenic memory, which is supposed to be an analogue of episodic memory. It was applied by Eacott and Norman [10] in a rat study; a change of open field background was introduced there instead of food locations’ change. Measuring of „when”, on the other hand, was excluded from the current implementation of the method. It has been repeatedly discussed before, and so far no direct connection between such „when” and „mental time travelling into the past”, which is a key feature of Tulving’s episodic memory concept [26], has been established. Even for a human adult, a recall of exact time of a past event or a chronological order of multiple past events may be much more difficult than recalling who was involved and what happened.

With regards to simultaneous measuring of autonoetic consciousness, this capacity has been already paired with episodic memory within developmental research, and usually 3-year-olds at youngest were involved [20]. Since verbal reports provide a measure of both episodic memory and autonoetic consciousness in these studies, they are designed for children who (1) have fairly developed language abilities, and (2) are able to verbalize personal experience [16][19]. Autonoetic consciousness derives from self-concept, and this self-concept is based on i. a. self-perception and self-recognition. Previous evidence suggests that self-perception is evinced by infants as young as 2-month-olds [7], whereas behavioural signs of self-recognition are traced back to the 6th month of life [6].

Gibson [13] introduced the term of visual proprioception, which points to the influence of vision on individual perception of self-movement beyond the interoceptive clues from the mechanical and vestibular system [7]. From the very beginning of their life, infant may perceive and interpret the features of surroundings based on visual proprioception. This way they can coordinate external clues and internal intentions [7]. Distinction between the self and the surroundings develops with the advancement of cognitive capabilities, and, consequently, alters the response to own and other’s images, both mirror reflections and videos. Since self-concept stems from self-perception, confrontation with mirror and videos in first 24 months of life may provide some insight into the self-concept development [7, p. 104]. After the 12th month of life a child may be able to find others via the mirror reflection and differentiate other videos from the videos of self. After the 15th month of life they may recognize her own face and recognize herself in the mirror. Finally, the 18th month of life brings unambiguous evidence of these abilities, for the signs of embarrassment towards own images emerge [15][22]. The onset of such self-recognition is only a step in the development of self-concept and autonoetic consciousness, and is identical with neither of these [21].

However, previous evidence points suggests that self-recognition may be evinced in behaviour long before the 12^{th} month of life. Fischer (1978), for instance, proposed a theory of cognitive development and actually tested it with a scale of tasks, which required different levels of distinction between self, surroundings, and others. With regards to this scale, it seems that self-recognition may not be about recognizing one's face, but rather differentiating oneself from the environment. This happens in the course of self-movement and coordination of three sets of clues: these from interoception ("what happens inside me", visual proprioception ("how I am moving" based on "what I am seeing") and appearance of surroundings ("how things look like and change when I am moving"). Therefore, the infant does not need to know the look of their face to be, to some extent, aware of their existence. At early stages of development they rarely, if ever, see their own face, whereas on daily basis look at their moving limbs [22]. For now, face is not subject to self-recognition. On the other hand, mirror self-recognition task is a widely accepted method of non-verbal measuring of self-recognition and self-awareness. In this task, briefly, a spot of rouge is placed on a participant's face, and then the participant is confronted with a mirror and expected to remove the rouge. Both a certain level of coordination and experience with self-images are prerequisite.

Method overview

Given the above, we aimed to design a method, which might offer, on one hand, some insight into the recollection of a personal past episode and, on the other, provide a simultaneous measure of behavioural signs of self-recognition, to the extent possible at the participant's age. Reliability of this method has been tested with a group of 17 children between the age of 13 and 16.5 months. This age range has been selected for three reasons: 1. as we mentioned above, some 15-month-olds may already pass the mirror-mark task; 2. they also may find others via mirror reflection and differentiate videos of self and others; 3. they are able to form strong and well organized memories of one-time events and recall them after a month, as evidence within elicited/deferred imitation paradigm has showed previously [4][5].

We adopted a solution proposed previously for the purposes of cognitive zoology research, to measure memory for "what"-who"-"where" features of the past event through differentiation of video recordings of this very event. We manipulated "what" via the toy alteration, "who" via the change of experimenter and "where" via the context change. We skipped the aspect of when for the reasons mentioned in the previous section.

The procedure consisted of two sessions with 24-hour delay (±3) in between. In general, on the first day, the child got familiar with the experimental space and the experimenter, participated in the mirror-mark task and then may have played with the experimenter a little. The session was video-recorded. On the second visit the child, sitting on the caretaker's knees, was watching the original and modified videos of relevant fragments from the initial visit.

Throughout the first session, the participant may have watched herself on a screen (real-time video of the on-going situation is being displayed on the screen within the child's view scope). Her interest in this image was recorded, so that we could compare the behavioural signs she evinced meanwhile with the behavioural signs evinced during the mirror-mark task, which follows. This was merely a pilot, which may be developed in the future. The behavioural signs included: touching the image accompanied with a vocalization, answering the "who's that" question with the child's version of the appropriate pronoun of her name, looking at the mother's reflected image and right back at her face, touching the face parts near the cream spot or touching the spot [1][6], and making faces toward the image. Making faces by children could be a social behavior towards the reflection, not a sign of self-recognition. However, it seems that children around the 15th month of life do not, for instance, poke their tongues at the others. These individuals among great apes who are to pass the mirror-mark test, initially pull their tongues out, pick their teeth, noses etc. [14, p. 86]. If the child smiled or cried, one may infer that she saw just a peer individual in the mirror, but poking tongue, pretending to make sounds in a soundless way, nose picking, legs stamping or „enacting an Indian" (making a noise with mouth and an open hand) do not belong to social behaviors at this age in the encounter with a 15-month-old complete stranger [12].

A crucial phase of the first session followed the real-time video and mirror-mark tasks. The participant was sitting on the caretakers' lap and the caretaker was sitting on a chair, while the experimenter, kneeling in front of them, set a toy in motion; either a land toy (a mouse and a caterpillar, each wind-up and with wheels) which was set a few times and rode on the ground, or a water toy (a seal and a dolphin, both wind-up and with fins) which, alternatively, was set a few times and swims in a meter long plastic box filled with water. The participant may have watched this play for a minute, and when the minute passed, she could join the experimenter. This play was recorded for future modification. At the end of the session, the participant was being rewarded and the visit was over.

During the 24-hour delay, the original 1-minute video recording was modified with Lightworks software for Linux. First 30 seconds were extracted and modified thrice, in terms of: 1. "what" - a land toy was substituted with a water toy and vice-versa; 2. "who" – a female experimenter is substituted with a male experimenter; 3. "where" – a violet curled up blanket is added on the ground, along the toy motion area or the box with water. The original recording was performed in such a way that merging it with three additional videos (1. with another toy, the same experimenter, 2. with the same toy, another experimenter, and 3. with the same toy, the same experimenter, with the blanket added) was possible with no (or almost in-) visible distortions. Therefore, the left part of the video (with the participant and her caretaker) was the same for all videos, and the right part came from one of these three additional videos, which were recorded in close time proximity to the original recording. Once all videos were produced in Lightworks, a test video was produced in Openshot software for Linux. Six thirty-second sets were composed, each of which included two videos: one original and one modified, with both clips displayed concurrently, one on the left side, and the second one on the right side of the screen (each video took 50% of the corresponding side, with black background equally distributed above and below). The test video was organized according to a fixed schema (see Figure 1).

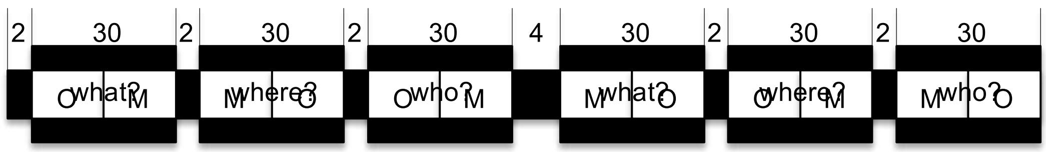


Figure 1. Test video schema. The numbers show duration in seconds. "O" and "M" stand for "original" and "modified" video. The order of "what"-"where"-"who" was individualized and random.

There were two sets with modified "what" aspect, two sets with modified "who" aspect and two sets with modified "where", one with the modified video on the left and original video on the right side of the screen, and one with reversed sides. A 1- second attention-grabbing sound ("meow" in this case) was played both during the 2-second within-set black screens and during the 4-second between-sets black screen

The participant was watching these videos on the second day, sitting on the caretaker's lap. The six sets of videos were displayed alternately. The order of "who"-"what"-"where" modified sets, and the side of the modified video in the first set were randomized for each child, and the side of the modified video was then systematically altered in each consecutive set. Each clip in each set was once presented on the left side and once on the right side of the screen. During the test the participant's face was recorded with the camcorder placed on the top of the screen for further analysis.

During the second session all behavioural signs of remembering the previous event were noticed by experimenter. They included the following acts: looking for the bubbles behind the table (where they fell on the previous day), trying to turn the screen on with the camcorder pilot (which worked on the previous day), looking for the camcorder, direct contact with the experimenter (in case of children who were distant on the previous day), searching for the cream and playing with the can (which was used on the previous day and remained in the same location). Such behaviors may provide an auxiliary measure of episodic recognition.

Looking times were used to assess participant's performance in the test. We recorded and analysed frame-by-frame the time of looking either at the left or the right part of the screen in Avidemux software for Linux. Precise beginning and ending of each video was registered in the mirror behind the participant. We judged eyesight direction based on eyes direction, and classified it into three categories: (1) left part of the screen; (2) right part of the screen, and (3) out of screen. The looking time was expressed in the number of watched frames. Whilst eyes were directed towards the left side of the screen, we counted the look at the right video, and once eyes moved and directed towards the right side, we counted the look at the left video. In neither case we counted the time of saccadic eye movements in between. We calculated looking time for each video and determined the proportion of watched genuine video frames to all watched video frames for each set and aspect, and we used a corresponding algorithm for modified videos as a measure of preference for a new event.

Results

We performed statistical analysis of the results with GNU PSPP software for Linux and ran single-sample t-tests for "what", "who" and "where". As expected, there was a significant difference in the scores for watching the genuine "what" videos ($M=0.41$, $SD=0.17$) than the modified "what" videos ($M=0.59$, $SD=0.17$), and proportion of watching time for modified video was significantly above 0,5; $t(16)=2.09$, $p=0.05$. There also was a significant difference in the scores for watching the genuine "who" videos ($M=0.4$, $SD=0.15$) and the modified "who" videos ($M=0.6$, $SD=0.15$), and proportion of watching time for modified video was significantly above 0,5; $t(16)=2.85$, $p=0.01$. Contrary to the expectations, there was no significant difference in the scores for watching the genuine "where" videos ($M=0.54$, $SD=0.12$) and the modified "where" videos ($M=0.46$, $SD=0.12$), and proportion of watching time for modified video was not significantly above 0,5; $t(16)=1,37$, $p=0.19$.

2x2 Table of Cross-Categorized Frequency Data test was performed via VassarStats to compare the effect of sibling possession, behavioral signs of recall and signs of self-recognition both in the mirror task and real-time video task on the overall time of watching the modified video in the aspects of what, who and where. No significant effects were confirmed due to the small number of participants; especially those who evinced mirror self-recognition.

Discussion of the method

Here, we aimed to address two cognitive capacities: "what"-"who"-"where" memory and self-recognition, and measure them simultaneously. As expected, children watched significantly longer the modified videos for "what" and "who" aspects of the previous experience than the genuine ones. This relationship was not so clear for the aspect of "where". Consequently, the "what" and "who" aspects of the previous experience were successfully recalled after the 24-hour delay, but evidence for recall of the "where" aspect was not obtained.

Although the insignificant result for "where" was not initially expected, it proved that the method was successful in terms of "who" and "what": participants' performance was a result neither of habituation to the genuine video nor of the eventual distortions which came from the computer modification (the genuine video was displayed alongside each modified film, and totally it was shown six times in a row, which may cause a gradual habituation, but its effect should be evident in the case of all three aspects). The lack of recognition of the "where" clues may have resulted from the fact that either the change of the context may not have mattered to the child or it was just not salient enough. This issue could be solved by a future computer manipulation of context as a whole, which was impossible in this case due to the technical issues. It could include a change of colour of the surrounding, alternation/addition of some items etc.

Introduction of both real-time video task and the mirror-mark task led to a few observations. Even those children, who were not preoccupied with their mirror image, were often interested in their video images. They made attempts of walking towards and away from the screen and expressed puzzlement when their head vanishes and only its very tip remained on screen. However, it seemed that experience with real-time videos may have affected this reaction, just like it happened in case of mirror-mark task. For instance, 14-month-old girl was particularly interested in her image, especially on screen: made faces towards, pointed at the image and answered her mother's „who's that?” question in the same way she named herself. Verbal cues were accepted in this task as a part of a natural interaction, advised in this case. Parents reported that approximately a month earlier their daughter had found out how to turn on the front camera mode on their mobile phones and had been recording videos of herself ever since. In the laboratory space she learnt to turn on and off the camcorder and in between the video display on screen she waited for the image. Once it appeared, she made faces and picked her nose again to make the adults laugh.

Self-recognition, however, was not crucial for the success in our task, which means that our method may not only work for children who already recognize themselves in the mirror, but also may be appropriate for those who do not evince this capability yet. As a result, it may serve as means of simultaneous measuring of self-recognition and episodic memory. Video recordings of the past event could also substitute the mirror-mark test in further research, as they refer to the self-movement patterns, and their distorted perspective of participant's actions may be more attractive for the participant herself. The success of real-time video recordings may also suggest that the episode is rather remembered as a whole, not a set of individual objects. If they memorized separate images of the surroundings, they would probably need more time to process genuine and modified videos, and, according to our results, children are able to differentiate them within thirty seconds (in each aspect).

However, our method, which received a draft name of altered video task, did not exploit an interesting opportunity here: to alter the partcipant-caretaker part of the recording. Due to the pilot character of the current study, we decided to focus on “what”-“who”-“where” aspects of the past situation (the significance of results was questioned during the Measuring Behaviour 2014 conference). Since the significant result was obtained for “what” and “who”, the further alterations could be introduced in future studies. For instance, the substitution of the participant could show, if she differentiated the videos of the actual event from these, where all “what”-“who”-“where” aspects remained unchanged, and the participant herself was altered. This could point to truly simultaneous recall of the past event and self-recognition.

The delay between the initial exposure and the test may be extended in the future, which would allow to trace the temporal boundaries of mental time travel at early developmental stages. Since episodic memory is about integration of the individual aspects of the past situation and their separate measurement proved successful, in the future procedures, the modifications may be merged into one video and set together with the genuine one.

This method may be tested in the future with younger children and, possibly, also with non-human animal species, at least those, whose vision would allow the introduction of such a procedure. Probably, as animals have generally less experience in watching videos than human infants, they will require longer exposure to camcorder-screen set and they may require an adequate adaptation of Fischer's five tasks [6]. Their environment taken into consideration, the setting would also require adequate changes and the range of possibilities in the aspect of “what”-”who”-”where” would be significantly wider. “What” could employ the use of different snacks or toys, “who” could either engage the presence of experimenter or conspecific, and “where” could involve the context change.

In conclusion, our method may be of use both within the developmental and the cognitive zoology approach. It still needs a series of improvements, but here we aim to test it with a group of participants, who could pass it, based on previous experience. It worked well for “what” and “who” aspects, and did not work for “where”, and we hope to discuss both these results with the Measuring Behavior audience.

References

1. Bard, K. A., Todd, B. K., Bernier, C., Love, J., & Leavens, D. A. (2006). Self-Awareness in Human and Chimpanzee Infants: What Is Measured and What Is Meant by the Mark and Mirror Test? Infancy 9 (2), 191-219.

2. Bauer, P. J. (2006). Constructing a past in infancy: a neuro-developmental account. Trends in Cognitive Sciences 10 (4), 175-181.
3. Bauer, P. J. (2007). Recall in Infancy. Current Directions in Psychological Science 16 (3), 142-146.
4. Bauer, P. J., & Leventon, J. S. (2013). Memory for One-Time Experiences in the Second Year of Life: implications for the Status of Episodic Memory. Infancy 18 (5), 755-781.
5. Bauer, P. J., Wenner, J. A., Dropik, P. L., Wewerka, S. S., & Howe, M. L. (2000). Parameters of Remembering and Forgetting in the Transition from Infancy to Early Childhood. Monographs of the Society for Research in Child Development 65 (4), 1-213.
6. Berthenthal, B. L., & Fischer, K. (1978). Development of Self-Recognition in the Infant. Developmental Psychology 14 (1), 44-50.
7. Butterworth, G. (1992). Origins of Self-Perception in Infancy. Psychological Inquiry , 3 (2), 103-111.
8. Carver, L. J., & Bauer, P. J. (1999). When the Event is More than the Sum of its Parts: 9-month-olds' Long-term Ordered Recall. MEMORY 7 (2), 147-174.
9. Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. Nature 397 (17), 269-274.
10. Eacott, M. J., & Norman, G. (2004). Integrated Memory for Object, Place, and Context in Rats: A Possible Model of Episodic-Like Memory? The Journal of Neuroscience 24 (8), 1948-1953.
11. Gaffan, D. (1994). Scene-specific memory for objects: A model of episodic memory impairment in monkeys with fornix transection. Journal of Cognitive Neuroscience 6 (4), 305-320.
12. Gazelle, H., & Faldowski, R. A. (2014). Peer Exclusion is Linked to Inhibition with Familiar but not Unfamiliar Peers at Two Years of Age. Infant and Child Development 23 (3), 220-228.
13. Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin.
14. Gallup, G. G. Jr. (1970). Chimpanzees: Self-Recognition. Science 167 (3914), 86-87.
15. Howe, M. L., Courage, M. L., & Edison, S. C. (2003). When autobiographical memory begins. Developmental Review 23, 471-494.
16. Hudson, J. A. (1990). The emergence of autobiographic memory in mother-child conversation. In R. Fivush, & J. A. Hudson (Eds.), Knowing and remembering in young children (pp. 166-196). New York: Cambridge University Press.
17. Lukowski, A. F., Wiebe, S. A., Haight, J. C., DeBoer, T., Nelson, C. A., & Bauer, P. J. (2005). Forming a stable memory representation in the first year of life: why imitation is more than child's play. Developmental Science 8 (3), 279-298.
18. Merchant, H., & Lafuente, V. d. (Eds.). (2014). Neurobiology of Interval Timing (Vol. 829). Springer.
19. Nelson, K. (1993). The psychological and social origins of autobiographical memory. Psychological Science 4 (1), 7-14
20. Perner, J., & Ruffman, T. (1995). Episodic Memory and Autonoetic Consciousness: Developmental Evidence and a Theory of Childhood Amnesia. Journal of Experimental Child Psychology 59, 516-548.
21. Povinelli, D. J., Landau, K. R., & Perilloux, H. K. (1996). Self-recognition in young children using delayed versus live feedback: Evidence for developmental asynchrony. Child Development 67, 1540-1544
22. Rochat, P. (2004). Origins of Self-Concept. In A. F. G. Bremner (Ed.), Blackwell Handbook of Infant Development. Wiley-Blackwell.
23. Schwartz, B. L., & Evans, S. (2001). Episodic Memory in Primates. American Journal of Primatology 55, 71-85.
24. Schwartz, B. L., Colon, M. R., Sanchez, I. C., Rodriguez, I. A., & Evans, S. (2002). Single-trial learning of "what" and "who" information in a gorilla (Gorilla gorilla gorilla): implications for episodic memory. Animal Cognition 5, 85-90.
25. Schwartz, B. L., Hoffman, M. L., & Evans, S. (2005). Episodic-like memory in a gorilla: A review and new findings. Learning and Motivation 36, 226-244.
26. Tulving, E. (2002). EPISODIC MEMORY: From Mind to Brain. Annual Reviews Psychology 53, 1-25.

Exploratory Analysis of Differential Physiological Responses to Emotional Film Clips

Y.Y. Tan[1,3], F. Tey[2] and B.D. Hatfield[1,3]

[1]Neuroscience and Cognitive Science Program, University of Maryland, College Park, U.S.A., ying2@umd.edu
[2]Combat Protection & Performance Program, DSO National Laboratories, Singapore, tliankhe@dso.org.sg
[3]Department of Kinesiology, University of Maryland, College Park, U.S.A., bhatfiel@umd.edu

1. Introduction

One of the many definitions of emotion describes the phenomenon as a set of coordinated physiological responses that prepares the organism for appropriate action and is triggered by salient environmental stimuli [1], [2]. Emotion is then seen as an adaptive response and creates an allostatic state to help in survival and daily functioning [3], [4], [5]. The evidence that distinct patterns of autonomic nervous system response exist for the six basic emotions, namely, happy, sad, fear, anger, surprise and disgust, was first presented by Ekman, Levenson and Friesen in 1983 [6], using "directed facial action" and "relived emotion". Since then, for the past thirty years, researchers across different domains of psychology and computer science have tried to uncover methods to differentiate and predict emotional states [7], [8], [9], [10], [11], [12].

The interest grew as the applications for being able to decode emotional states were far and wide. Other than to obtain greater evolutionary and theoretical understanding [13], [14], computer scientists viewed it as a way to improve artificial intelligence. By allowing machines to accurately identify users' emotional reactions in their interactions, it encourages users to become fonder of using the machine and develop a stronger attachment, thereby making the technology more successful [9], [10]. Psychological studies of human emotions, memory and decision making [15], [16], [17], [18] also prompt computer scientists to envision machine systems that might one day make use of emotion states to make more human-like and "intelligent" decisions. Others study emotions to understand how affective disorders through physiological disruptions and sustained arousal can affect physical health [19], [20], [21], [22], [23].

In this present study, we seek to explore the possibility of classifying physiological responses according to emotion conditions and examine whether physiological systems are recruited as a coherent whole or independently. In regard to the latter, Mauss and colleagues [24] have reported coherent responses between physiology and behavior, but to date, no known studies of correlating simultaneously activated responses between different physiological systems of the autonomic nervous system within a human body have been conducted. Therefore, the present study was conducted to determine the proposed coherent set of responses existing for different emotion states.

Methods

Participants

Sixty participants (30 males and 30 females, M = 19.4, SD = 1.25 years, range = 18 – 23 years) were recruited through polytechnics in Singapore via their teachers and the school's online announcement platform. As the measurement electrodes are placed on the participants' left hand, their hands were loosely restrained to prevent movement artefacts. Thus, we only recruited right-handed participants. They received SGD50 for their participation. They were excluded if they indicated that they had a history of or were suffering from cardiovascular disease, psychological or psychiatric problems. They were advised to abstain from caffeine and alcohol the day before the experiment. The study was approved by the Institutional Review Board of DSO National Laboratories, Singapore and University of Maryland, College Park.

Materials

Twelve film clips were presented to evoke the targeted basic emotions of happiness, sadness, fear, anger, surprise and disgust conditions and one neutral film was also selected to be played at the beginning of each experiment as the neutral condition. They were presented in six different pseudo-randomized orders. The orders were designed in a way in which no two film clips intended to elicit the same emotion states were presented consecutively and that the happy and sad conditions were not presented consecutively. Participants were randomly placed in one of these presentation orders. The film clips were selected and validated out of 42 films clips. They were highly rated and ranked by one hundred and thirty participants (63 males and 67 females, M = 19.0, SD = 1.44, range = 18 – 23 years) in group and individual settings. Most films were taken from Gross and Levenson's [25] and Schaefer and colleagues' [26] studies but included Asian films catered for the local population. The films ranged from 45 to 370 seconds and were presented on a 13.1" LCD monitor. Sounds from the clips were delivered through headphones to attenuate extraneous noise and allow participants to stop the sounds easily by taking the headphones off any time during the film clip if they felt distressed.

A post-film questionnaire modelled after Rottenberg, Ray and Gross's [27] suggestion was used in this study to evaluate the emotions felt while watching the clips. The questionnaire was administered on a laptop next to the monitor with a USB mouse attached. There were six emotions that participants have to rate even if they did not feel it. They were the six targeted emotions; happy, sad, fearful, angry, surprised and disgusted. Questions included the intensity of emotions felt on a scale from 0 – 8 (where 0 is *neutral* and 1 to 8 as the emotion increases in intensity), any other emotions they felt while watching the films and their perceived intensity based on the provided scale, whether they watched the film before and if they closed during the presentation. Participants were encouraged not to change their answers once they have entered it in the questionnaire.

Procedure

Experimental sessions were conducted in a sound attenuated room where the participants and experimenters were separated by a solid screen and participants were observed and recorded via a webcam installed on top of the monitor. Participants were seated comfortably in a stationary chair approximately 50 – 60 cm away from the screen. The top frame of the monitor was levelled with the participants' eyes. Their left hands were loosely restrained and rested in a comfortable position on a table next to their seat. The table was up to the participant's waist level when they were seated. They were told to refrain from moving when the clip was playing. Participants were instructed to pay close attention to their emotions while watching the clips so that they could accurately answer the post-film questionnaire. They were allowed to take the headphones off, close their eyes or look away at any time during the playing of the clip if they felt uncomfortable. After viewing each clip, they were prompted to enter their responses on a laptop next to the monitor. Before the experiment began and in between film clips, participants were given a 6.5 minutes "washout period" [12], where they copied excerpts from articles that described the history of transportation and vehicles. These passages were controlled to be as emotionally neutral as possible. This "washout period" was designed to help participants forget about the previous film and enable their physiological responses to return to baseline. Participants were informed about the intent of the "washout period" and were told to concentrate on the present film. Furthermore, they were told not to ruminate about the previous film clips they have watched. In between 2 – 3 clips, participants are given a one minute break to move their restrained hand (with sensors still attached) and move around in their seat.

Physiological measures and quantification of physiological data

Physiological measurements were recorded using the Procomp Infiniti Encoder (SA7500; Thought Technology Ltd., Montreal, CA) at 256 Hz and digitized using the accompanied software, BioGraph Infiniti's Physiology Suite. All quantified measures were input as features for analysis and modeling.

For pattern classification algorithms to work optimally, it is best that data values are close to one another across variables. Therefore, the arbitrary values of skin conductance and respiration were standardized using the below equation:

$$x_i = \frac{x_i - \overline{x}}{\sigma_x}$$

where x_i is the observed value, $\overline{x}$ is the mean of time series and σ_x is the standard deviation of the time series. For cardiovascular measures, since interbeat intervals (IBIs) have to be derived from the EKG measure, they were not standardized. The IBI values were not standardized as well as their values are important indices for heart activity. Moreover, their values ranged from 0 – 2 in our experiment, which was within the standardized values.

Electrocardiogram was obtained using the provided EKG-Flex/Pro sensor with Uni-Gel self-adhesive electrodes (T3425; Thought Technology Ltd., Montreal, CA) attached to the participants' thoracic areas in a Lead II configuration. Measurement sites were first prepared with 70% isopropyl alcohol. A peak detection algorithm was written in MATLAB to identify R peaks in the *PQRST* wave of an electrocardiogram (ECG). The times between two consecutive peaks are then computed as the IBI. Excessive short (less than 300 ms) or long beats (more than 2000 ms) were corrected by removal as artefacts or insertion as missed beats respectively [28]. The mean, variance, maximum and minimum values were obtained for the IBI and amplitudes for the fluctuations of the IBI. The root mean squared successive difference (RMSSD) of the IBI was also calculated as an index for heart rate variability [29].

Skin conductance was measured using the SC-Flex/Pro sensor with two Uni-Gel electrodes attached to the participants' left palm, specifically the thenar and hypothenar eminences. From the standardized values, the mean, variance, maximum and minimum values were obtained for the skin conductance levels, the amplitude of the fluctuations and the period of fluctuation. The number of peaks was also counted as an index of fluctuation.

Respiration was measured using the Respiration-Flex/Pro stretch sensor strapped around the participants' upper chest. The values were passed through a peak detection algorithm in MATLAB to obtain its amplitude, period and rate (respiration per minute). From the standardized values, the mean, variance, maximum and minimum values were obtained for the inspiration and expiration depth, the amplitude of the fluctuations and the period of fluctuation. The number of peaks was also counted to infer respiration rate.

Analysis

Film and case selection

Due to the fact that not all participants reported feeling the targeted emotion or that they reported feeling a blend of emotions, there was a need to extract data that exhibit strong characteristics of the emotion condition (i.e. targeted emotion rated highly) and is consistent across participants (i.e. participants do not show too much variability in ratings between them) so that the patterns of physiological reactions experienced for each emotion condition can be inferred with better confidence. Otherwise, there could be the possibility that the variability might be attributed to other extraneous factors other than feeling the targeted emotion.

Hence, instead of collapsing data across film in similar emotion conditions, only data from the more effective film per emotion condition was chosen. The surprise condition was not analyzed here as the film clips that were chosen could only elicit the targeted emotion towards the end of the clip. As a result, there were insufficient data points for analysis (i.e. less than 60 seconds). The anger condition was also not considered for analysis as the films were unable to elicit a reasonable level of the targeted emotion (mean of less than 4.0) and also elicited a comparable level of sadness as shown in Figure 10 below.

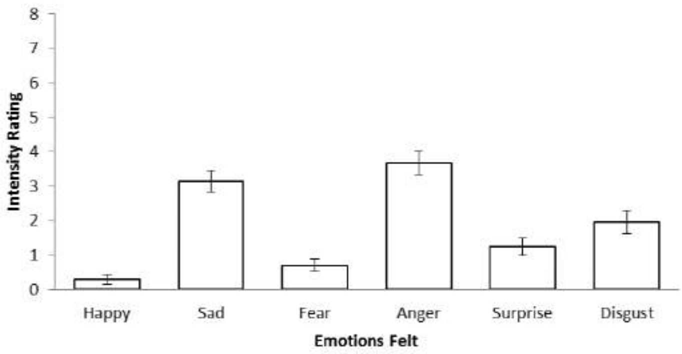


Figure 10. Self-reported ratings for film clip inducing anger obtained from post-film questionnaire. Whiskers represent 1 standard error of the mean.

Additionally, two criteria based on self-report ratings in the post-film questionnaire were established for case selection.

8. Participants must rate the targeted emotion of the film to be at least a '4'.

9. Participants must not rate other emotions other that the targeted emotion to be more than '2'.

Based on these criteria, only twenty participants qualified in the disgust condition. As such, for comparison purposes, participants for each condition were ranked according to the rating of the targeted emotion (where participants who gave the highest rating were ranked first) in the emotion condition and the first twenty cases were selected. The results before and after case selection are presented in Table below. When all participants were included, targeted emotions were rated statistically higher than non-targeted emotions (all $p < 0.05$). No statistically significant differences were found between randomization orders or genders. For the selected group of participants, the mean ratings for targeted emotions are higher than when all participants were included. Non-targeted emotions are also rated lower or were almost non-existent.

Table 1. Mean (SD) of self-report ratings for each emotion by emotion condition for all participants and selected group of participants.

All Participants	Emotion Condition									
	Neutral		Happy		Sad		Fear		Disgust	
Emotion Self-Report	M	SD	M	SD	M	SD	M	SD	M	SD
Happy	1.23	1.83	5.84	2.19	0.90	1.69	0.48	1.47	1.29	2.11
Sad	0.02	0.13	0.02	0.13	5.05	2.30	0.44	1.10	0.15	0.57
Fear	0.06	0.51	0.00	0.00	0.89	1.86	5.21	2.57	0.44	0.97
Anger	0.00	0.00	0.11	0.77	0.81	1.82	0.35	1.34	0.13	1.02
Surprise	1.29	1.95	0.79	1.54	0.26	0.89	2.21	2.53	1.89	1.99
Disgust	0.00	0.00	0.02	0.13	0.16	0.63	0.27	0.85	5.08	2.61
Selected	Emotion Condition									
	Neutral		Happy		Sad		Fear		Disgust	
Emotion Self-Report	M	SD	M	SD	M	SD	M	SD	M	SD
Happy	0.00	0.00	7.65	0.49	0.25	0.55	0.05	0.22	0.10	0.31
Sad	0.00	0.00	0.00	0.00	6.35	0.99	0.20	0.41	0.05	0.22
Fear	0.00	0.00	0.00	0.00	0.05	0.22	7.05	1.23	0.30	0.66
Anger	0.00	0.00	0.00	0.00	0.15	0.37	0.05	0.22	0.00	0.00
Surprise	0.00	0.00	0.00	0.00	0.05	0.22	0.60	1.23	0.55	0.76
Disgust	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.22	6.60	1.57
Gender	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
	12	7	8	12	11	9	15	5	14	6

Data extraction

Due to the nature of films, where emotions can fluctuate over time [24] or takes time to develop [11], it is important to capture the moment where the emotion experience is at its peak. It is also important for different emotion condition to have similar data bin lengths for comparison purposes. In order to achieve this, one minute intervals were extracted from each emotion condition. For neutral, disgust, happy and fear conditions, where films tend to be shorter because they are easier to elicit almost immediately, the first seventy seconds of data was considered and the first ten seconds of data was ignored. For the sad condition, where films tend to be longer as context has to be given [11] the last seventy seconds of data were extracted and the last ten seconds of data ignored.

Results

All statistical analysis was done using SPSS 21.0. This includes repeated measures ANOVA with Tukey honestly significant difference (HSD) post-hoc tests, factor analysis employing principal component extraction and discriminant analysis. Support vector machines modeling was done using the Statistical Toolbox in MATLAB 2013a. The functions include svmtrain and svmclassify.

Univariate analysis of the effects of emotions

Univariate ANOVA was conducted on all derived features to investigate if there are significant main effects between emotion conditions. Tukey honestly significant difference (HSD) post hoc test was conducted if significant differences were found.

The fear condition revealed significantly lower mean IBI (i.e. higher heart rate) than the sad condition (F=3.48, p<0.05) and higher variance (F=3.37, p<0.05) and number of peaks (F=3.61, p<0.05) for skin conductance levels than the neutral and sad conditions. This reflects a heighten state of arousal and reactivity to stimuli as seen in the heart and skin conductance response respectively that is often described to facilitate the escape or avoidance tendencies of fear (Ekman, 1999; Levenson, 1999, 2003).

The sad condition revealed significantly longer respiration periods (F=3.14, p<0.05) than the happy condition and lower variance (F=2.85, p<0.05) than the disgust condition. It also revealed significantly lower maximum skin conductance levels (F=3.24, p<0.05) and variance in respiration depth than the neutral condition (F=2.56, p<0.05). This is consistent with a depressed, withdrawn state that is described by Levenson (1999) that arises from the tendency to want to seek sympathy from other and remain passive rather than react actively.

The disgust condition revealed significantly higher variance of respiratory period than the sad condition (F=2.85, p<0.05) that could be caused by the gag and vomit reflex to expel toxins from the body (Purves et al., 2013). In the happy condition, participants were observed to be laughing when watching the film clip during the experiment. It was expected that as a result, a lower mean period of respiration would be produced. Indeed, it was found that the mean period of respiration in the happy condition was significantly lower than in the sad condition. Although there were no significant differences in the minimum period of respiration, the disgust and happy conditions displayed lower values than the other emotion conditions due to their automatic triggering of gag and laughter.

The neutral condition elicited higher values than all other emotion conditions for mean skin conductance levels and minimum skin conductance level. This could be attributed to the apprehension or excitement from participating in the study.

Factor analysis

This method was employed to investigate if different features for different emotion conditions can be grouped according to their emotion condition and if different emotion conditions can be considered different from one another based on these features. More importantly, highly correlated features grouped under a single factor can also inform us about what kind of physiological responses are occurring in synchrony and which emotion condition would bring about this coherent state.

As shown in Table 2, most of the features were grouped under each component according to their emotion conditions. The fear and sad conditions accounted for most of the variance in the factor analysis. No emotion conditions, other than the fear condition, had both cardiovascular features and skin conductance features grouped under one component (i.e. *Fear 1*). This could indicate a coordinated response in two physiological systems that could be brought about only by fear. The components named *Sad1*, *Sad 2* and *Sad 3* show the three different physiological systems measured grouped under three different components. There were three components for fear as well, but *Fear 2* included only skin conductance features and other skin conductance features were already grouped under *Fear 1*. *Fear 3* represented distinctly respiration features for the fear condition. Thus, the fear and sad conditions seems to be very well represented by the three physiological systems measured as compared to the other emotion conditions.

Table 2. Results from factor analysis. Abbreviations: measures are in caps - HR: heart rate, SC: skin conductance, R: respiration; statistical derivatives of measures are placed before the measure – m: mean, v: variance, min: minimum, max: maximum; characteristics of statistical derivatives over time are placed after the measure – a: amplitude, p: period; emotion conditions are placed at the end of the abbreviations separated by a space – f : fear, s: sad, d: disgust, h: happy, n: neutral.

Fear 1		Sad 1		Sad 2		Happy 1		Disgust		Neutral 1		Fear 2		Sad 3		Neutral 2		Fear 3	
mHRa f	0.98	minHRa s	0.96	maxRa s	0.90	rmssd h	0.96	minSCa d	0.94	rmssd n	0.94	mSC f	0.90	mSCp s	0.91	minHRa n	0.91	vR f	-0.93
minHRa f	0.98	mHRa s	0.94	vR s	0.87	vHRa h	0.93	minSCp d	0.89	maxHRp n	0.87	mSCa f	0.83	maxSCp s	0.88	mRp n	0.82	maxRa f	-0.91
vHRp f	0.98	vHRp s	0.94	vRa s	0.84	mHRa h	0.93	minSCa n	0.86	maxHRa n	0.86	maxSC f	0.80	minSCp s	0.88	nR n	-0.75	maxR f	-0.91
vHRa f	0.98	rmssd s	0.89	maxR s	0.81	maxHRa h	0.91	mSCp d	0.83	vHRa n	0.85	nSC f*	0.77	vSCa n	0.74	mHRa n*	0.73	vRa f	-0.82
rmssd f	0.96	mHRp s	0.80			maxHRp h	0.90	mSCa n	0.81	vHRp n*	0.74	maxSCa f*	0.76						
mHRp f	0.92	minHRp s	0.72			vHRp h	0.86	mSCa d	0.79			minSCa f	0.74						
maxHRp f	0.90					mHRp n	0.82					minSC f	0.74						
vSCp f	0.90																		
maxHRa f	0.87																		
vSCa f	0.81																		
minHRp f	0.72																		
mSCp f	0.71																		
mHRa d	0.88																		
mHRp d	0.80																		
vHRp d	0.76																		
minR h	-0.85																		
maxRp n	0.72																		
Eigenvalue	22.73	Eigenvalue	20.36	Eigenvalue	18.25	Eigenvalue	14.50	Eigenvalue	12.19	Eigenvalue	11.00	Eigenvalue	9.36	Eigenvalue	8.50	Eigenvalue	8.44	Eigenvalue	7.52
% Total Variance	12.99	% Total Variance	11.63	% Total Variance	10.43	% Total Variance	8.29	% Total Variance	6.97	% Total Variance	6.28	% Total Variance	5.35	% Total Variance	4.86	% Total Variance	4.82	% Total Variance	4.30
Cumulative Variance (%)	12.99	Cumulative Variance	24.62	Cumulative Variance	35.05	Cumulative Variance	43.34	Cumulative Variance	50.30	Cumulative Variance	56.59	Cumulative Variance	61.94	Cumulative Variance	66.79	Cumulative Variance	71.62	Cumulative Variance	75.91

Note. Extracted using Principal Component Extraction, rotated using Varimax with Kaizer normalization. Only variables with loadings more than 0.70 are shown. *also loaded on other components.

4.3. Discriminant analysis

In order to distinguish between different emotion conditions and to see if it is possible to model the different emotion conditions, discriminant analysis was carried out on all features. The results are presented in Table 3. Overall, 86% of cases were classified correctly.

Table 3. Discriminant analysis results for all emotion conditions

		Predicted Group Membership					
		Neutral	Disgust	Happy	Fear	Sad	Total
Original	Neutral	17 (85%)	0 (0%)	1 (5%)	1 (5%)	1 (5%)	20
	Disgust	0 (0%)	19 (95%)	1 (5%)	0 (0%)	0 (0%)	20
	Happy	3 (15%)	2 (10%)	13 (65%)	1 (5%)	1 (5%)	20
	Fear	0 (0%)	0 (0%)	0 (0%)	18 (90%)	2 (10%)	20
	Sad	1 (5%)	0 (0%)	0 (0%)	0 (0%)	19 (95%)	20
	Total	21	21	15	20	23	100
86% of cases correctly classified							

As revealed by ANOVA and factor analysis, fear and sad conditions were physiologically distinct from many of the other emotion conditions. When the classification analysis was ran with only those two conditions and the neutral condition, the classification improved by 10.7% as shown in Table 4.

Table 4. Discriminant analysis results with only neutral, fear and sad conditions.

		Predicted Group Membership			
		Neutral	Fear	Sad	Total
Original	Neutral	19 (95%)	0 (0%)	1 (5%)	20
	Fear	0 (0%)	19 (95%)	1 (5%)	20
	Sad	0 (0%)	0 (0%)	20 (100%)	20
	Total	19	19	22	60
96.7% of cases correctly classified					

In summary, discriminant analysis performed well in modeling existing data, however, when trying to fit new, unfamiliar cases it performed poorly (i.e. only 40 – 60% correct classifications). For future prediction purposes, it is necessary to find another method to classify the data. While discriminant analysis tries to find a linear relationship between features and places new cases according to probabilistic models, support vector machine (SVM) identifies patterns between data points in the same conditions and simply finds a hyperplane that can separate data points from different conditions in a multidimensional space. The hyperplane that induces the least cost is chosen.

The MATLAB functions svmtrain and svmclassify were used for binary classification and the emotion conditions fear and sad were chosen for analysis. This stems from the fact that these two conditions have shown considerable differences from other conditions and good component representation results thus far. The default settings of sequential minimal optimization method and a 3rd order polynomial linear kernel function were kept.

4.4. Support Vector Machine (SVM)

The existing data set for fear and sad conditions were used as the training set. Two test sets were used for validation. The first test set was derived from selecting the next five participants down in the earlier ranking done for the first twenty participants used in the training set. Based on the case selection criteria described earlier, the top five participants viewing the other film for each of the emotion condition were chosen as the second test set. For the first test set, SVM only managed to group 40% of the cases (i.e. 2 out of 5) correctly for the fear condition and 60% of the cases (i.e. 3 out of 5) correctly for the sad condition. However, for the second test set, SVM managed to group all new cases correctly with 100% success. This result hints at a possible threshold level for subjective feeling for it to have an effect on the autonomic nervous system.

5. Discussion

The results revealed that the fear and sad conditions were distinct from one another and from baseline. The separation from and between the other emotion conditions were not so clear cut, but this was to be expected. Kreibig and colleagues [11] recognized the “psychological, experiential, expressive, and behavioral distinctions between fear and sadness” and set out to only try to distinguish between these two emotion conditions with relatively high success rates of 69 to 84.5% success rates in their classification analysis. For this study, successful classification rates were higher; from 96.7% up to even 100%. The elevated success could be due to the selection of participants according to how successful the film clip elicited the targeted emotion, which gave a more accurate representation of the resultant autonomic nervous system activity.

The results from SVM also showed that for emotion conditions to have an effect on the autonomic nervous system, participants will first have to feel a certain level of the emotion. This could be a high intensity of 7 – 8 as represented by the second test set. It could also be that the model was trained using data from participants that rated the targeted emotions highly. However, discriminant analysis showed that this would yield effective differentiation between emotion conditions and so, this should not be a disadvantage.

It was unfortunate that the neutral condition was unable to create a representative baseline state for comparisons with other emotion conditions as it has shown significantly higher skin conductance level, indicating a high level of arousal that could have been a product of the excitement from starting the experiment. If the study were to be repeated, it might be wise to randomize the presentation order of the neutral film clip as well. Instead of presenting it first all the time, like in this study, perhaps it would be advisable to insert it in between other films as well to prevent experiment novelty from triggering high skin conductance levels.

In trying to investigate the coherence between different physiological systems, factor analysis yielded correlations between cardiovascular response and skin conductance in the fear condition only. The other emotion conditions, including the sad condition, showed that physiological systems might be working relatively independently from one another, causing perhaps an incoherent activity state. This could stem from the fact that the fear response is for fight-or-flight, and it has been the most common emotion that has been described with the most detail about the consequent coordinated responses from different systems (for example, increase in heart rate, increase in skin conductance, increase in respiration) [11], [12], [13].

In the field of cognitive neuroscience, synchronicity and coherence in the brain has been found to improve neural efficiency and cognitive performance [30], [31], [32]. It will be of interest to investigate if the same principle applies to the autonomic nervous system as well. From a more global perspective, maybe incorporating a more holistic system and interaction between the brain and the body can give better predictions and explanations regarding emotion states and its effect on cognitive performance.

In summary, the analysis and results were used to show that it is important to select participants that exhibit a high level of a targeted emotion for classification and prediction of emotion states. A successful elicitation and classification can then be used in other applications and applied with more confidence. It may be that a large number of data will be excluded, but it is pertinent for a good conclusion.

References

1. P.R. Kleinginna Jr, and A.M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition", Motivation and emotion, vol. 5(4), 1981, pp. 345-379.
2. N.H. Frijda, “Impulsive action and motivation”, Biological psychology, vol. 84(3), 2010, pp. 570-579.
3. N.H. Frijda, “The laws of emotion”, American psychologist, vol. 43(5), 1988, pp. 349.
4. R.W. Levenson, “Autonomic nervous system differences among emotions”, Psychological science, 1992, vol. 3(1), pp. 23-27.
5. M.M. Bradley and P.J. Lang, “Measuring emotion: Behavior, feeling, and physiology”, Cognitive neuroscience of emotion, 2000, vol. 25, pp. 49-59.
6. P. Ekman, R.W. Levenson and W.V. Friesen, “Autonomic nervous system activity distinguishes among emotions”, Science, 1983, vol. 221, pp. 1208-1210.
7. R.W. Levenson, P. Ekman and W.V. Friesen, “Voluntary facial action generates emotion-specific autonomic nervous system activity”, Psychophysiology, 1990, vol. 27, pp. 363-384.
8. P.J. Lang, M.K. Greenwald, M.M. Bradley and A.O. Hamm, “Looking at pictures: Affective, facial, visceral, and behavioral reactions”, Psychophysiology, 1993, vol. 30, pp. 261-273.
9. R.W. Picard, E. Vyzas and J. Healey, “Toward machine emotional intelligence: Analysis of affective physiological state”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, vol. 23, pp. 1175-1191.
10. K.H. Kim, S.W. Bang and S.R. Kim, “Emotion recognition system using short-term monitoring of physiological signals”, Medical and biological engineering and computing, 2004, vol. *42*, pp. 419-427.
11. S.D. Kreibig, F.H. Wilhelm, W.T. Roth and J.J. Gross, “Cardiovascular, electrodermal, and respiratory response patterns to fear-and sadness-inducing films”, Psychophysiology, 2007, vol. 44, pp. 787-806.
12. C.L. Stephens, I.C. Christie, and B.H. Friedman, “Autonomic specificity of basic emotions: evidence from pattern classification and cluster analysis” Biological psychology, 2010, vol. 84, pp. 463-473.
13. P. Ekman, “Basic emotions”, Handbook of cognition and emotion, 1999, vol. 98, pp. 45-60.
14. R.W. Levenson, “The intrapersonal functions of emotion’, Cognition & Emotion, 1999, vol. 13, pp. 481-504.

15. A.R. Damasio, B.J. Everitt, and D. Bishop, "The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]", Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 1996, vol. 351, pp. 1413-1420.
16. E.A. Phelps, "Human emotion and memory: interactions of the amygdala and hippocampal complex", Current opinion in neurobiology, 2004, vol. 14, pp. 198-202.
17. E.A. Phelps, "Emotion and cognition: insights from studies of the human amygdala", Annu. Rev. Psychol., 2006, vol. 57, pp. 27-53.
18. E.A. Kensinger, "Negative emotion enhances memory accuracy behavioral and neuroimaging evidence", Current Directions in Psychological Science, 2007, vol. 16, pp. 213-218.
19. J.J. Gross and R.W. Levenson, "Hiding feelings: the acute effects of inhibiting negative and positive emotion", Journal of abnormal psychology, 1997, vol. 106, pp. 95-103.
20. J.J. Gross, "Antecedent-and response-focused emotion regulation: divergent consequences for experience, expression, and physiology", Journal of personality and social psychology, 1998, vol. 74, pp. 224-237.
21. R.M. Carney, J.A. Blumenthal, P.K. Stein, L. Watkins, D. Catellier, L.F. Berkman, S.M. Czajkowski, C. O'Connor, P.H. Stone and K.E. Freedland, "Depression, heart rate variability, and acute myocardial infarction" Circulation, 2001, vol. 104, pp. 2024-2028.
22. R.M. Carney, K.E. Freedland, and R.C. Veith, "Depression, the autonomic nervous system, and coronary heart disease" Psychosomatic medicine, 2005, vol. 67(Supplement 1), pp. S29-S33.
23. J. Rottenberg, and J.J. Gross, "When emotion goes wrong: Realizing the promise of affective science", Clinical Psychology: Science and Practice, 2003, vol. 10, pp. 227-232.
24. I.B. Mauss, R.W. Levenson, L. McCarter, F.H. Wilhelm, and J.J. Gross, "The tie that binds? Coherence among emotion experience, behavior, and physiology" Emotion, 2005, vol. 5, pp. 175-190.
25. J.J Gross, and R.W. Levenson, "Emotion elicitation using films". *Cognition & Emotion*, 1995, vol. *9*, pp. 87-108.
26. A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers", Cognition and Emotion, 2010, vol. 24, pp. 1153-1172.
27. J. Rottenberg, R.D. Ray, and J.J. Gross, "Emotion elicitation using films". In J. A. Coan &. J. J. B. Allen (Eds.), The handbook of emotion elicitation and assessment, 2007 London: Oxford University Press.
28. G.G. Berntson, K.S. Quigley, J.F. Jang, and S.T. Boysen, "An approach to artifact identification: Application to heart period data", Psychophysiology, 1990, vol. 27, pp. 586-598.
29. Task Force of the European Society of Cardiology, the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation and clinical use", Circulation, 1990, vol. 93, pp. 1043-1065.
30. M.J. Hogan, G.R. Swanwick, J. Kaiser, M. Rowan, and B. Lawlor, "Memory-related EEG power and coherence reductions in mild Alzheimer's disease", International Journal of Psychophysiology, 2003, 49(2), 147-163.
31. J.C. Rietschel, M.W. Miller, R.J. Gentili, R.N. Goodman, C.G. McDonald, and B.D. Hatfield, "Cerebral-cortical networking and activation increase as a function of cognitive-motor task difficulty" Biological psychology, 2012, vol. 90, pp. 127-133.
32. P.J. Uhlhaas, and W. Singer, "Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology" Neuron, 2006, vol. 52, pp. 155-168.

Baby FaceReader: Designing an optimal model to track and model infant faces

A. Maroulis and M. den Uyl

VicarVision, Amsterdam, the Netherlands. andreas@vicarvision.nl

Abstract

Although automatic facial coding has received much attention in the past decade, it has largely focused on the modeling and coding of adult faces. Facial coding of infant faces is still under developed as it is limited by the variability of faces used in previous attempts. This paper evaluates 5 artifically built face models using Active Appearance Models against 2 datasets of images with baby faces. Future directions are discussed.

Introduction

Coding human behavior is an often-employed practice in understanding emotional and cognitive processes as well as describing social interaction. Comprehensive manual behavior coding processes, however, are quite cumbersome, time-consuming and, consequently, costly. Advances in the field of computer vision have led to breakthroughs in automating these coding processes. With relatively high-accuracy one can unobtrusively track and describe movement in the human body [1], the eyes [2] as well as the face [3] consistently, fast and on a large scale.

Infant (ages 0-2 years) facial coding has been present since 1920 [4] and is used to help answer research questions related to emotional, cognitive and social development, since infants lack the ability to communicate in a coherent verbal manner. In more recent years, researchers have looked to the infant face to identify genuinely felt pain [5, 6], evaluate liking and preference in eating [7], and describe social communication during typical and atypical development [8]. Although present for nearly a century, however, there has been little effort to automate the process of infant facial coding.

Most efforts to automatically track and code the human face are based on models of adult faces that do not generalize well to faces of all age groups. Infant faces vary from adult faces in that they are smaller, rounder, have a considerable amount of subcutaneous fat, elastic skin and often have little to no eyebrows [9]. To deal with these physiological differences there exist facial coding manuals in the literature tailored to infant faces [5, 7, 10]. Similarly, in the field of computer vision, the limited attempts in automated facial tracking and coding of infant faces are specific to that age group [8, 11-13]. These attempts, however, are either built using a limited number of faces in a laboratory environment (e.g. 12 faces) their model is based on [8, 11] or use modeling techniques that do not generalize well to varying lighting and angle [12, 13]. Most automated facial coding solutions for adult faces build their models on numbers in the scale of hundreds collected both (e.g. Noldus' FaceReader [14], Affectiva's Affdex). It is thus a natural next step to apply the same variance in faces in building a model for baby faces using a great number of images of baby faces to make it generalizable to as many infant faces as possible.

In this paper, we evaluate 5 artificially synthesized face models (texture and shape) on 2 test sets of images of baby faces (40 different babies each). The face models are built using the Active Appearance Models (AAM) method [15]. The AAM method uses a set of landmark annotated images to identify sources of variance amongst images. It then uses Principal Component Analysis (PCA) to reduce the dimensionality of the model. In other words, PCA is used to reduce the number of variables that are used to express the model, identifying what variables are most representative in the model. After the mean synthesized model is created from the annotated images, each new face image introduced is expressed as deviation of that mean face. The face models we compare are the following:

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

10. The *Noldus FaceReader General Model* [14]: a commercially available solution designed to identify and model adults' faces

11. The *Noldus FaceReader Children Model* [14]: a commercially available solution designed to identify and model children's faces (ages 2-8 years)

12. A *Google Images Baby Model:* a custom built model made up of images collected from several searches on Google Images for baby faces (estimated ages 0-2 years). This is designed to accommodate videos and images of baby faces collected "in the wild".

13. An *Experiment Baby Model:* a custom built model made up of images collected from several university studies on infant faces using Face-to-Face Still face (FtF) paradigm and free Parent Child Interaction (PCI) (actual ages 0-2 years). This is designed to accommodate videos and images of baby faces collected during experimental studies.

14. A *Combined Baby Model:* a custom built model made up of images in the Google Images Baby Model and the Experimental Baby Model, images of young children and few adults. This is designed to accommodate videos and images of baby faces collected in any setting.

We will compare the performance of all 5 models against 2 datasets:

1. *Google Images Baby Test Set:* a set of 40 baby face images collected from Google image searches of baby faces that have not been used in creating the *Google Images Baby Model*.

2. *Experimental Baby Test set*: a set of 40 baby face images collected from the same datasets used to build the *Experimental Baby Model* that have not been used to create the *Experimental Baby Model.*

Image and dataset requirements

For the purposes of this paper we have set the following requirements for all images included in the proposed models and test sets:

1. Each image has only 1 baby face visible (ages 0 – 2 years old)

2. The baby face must be facing front or at a viewing angle of up to 15^o of up from an observer's point of view

3. The full face of the baby must not be occluded in any way (eg. an object or the baby's hands are blocking any are from the chin to the forehead

4. The face must be lit with uniform lighting

A separate requirement is only one image per baby at a specific age can be used by either the model image sets or the test set. This way, we can ensure that the fit of a model is not affected by what images it is being tested on.

Evaluation Method

Each of the 5 models will be evaluated on both the *Google Image Baby Test Set* and the *Experimental Baby Test Set*. The metrics that the models will be evaluated on are:

- *Percentage of faces modeled*: A higher percentage indicates a better model.

- *Fitting Error* (as used in [16]): a lower score indicates a better model fit.

The authors will also review evaluation methods proposed by [17] and consider including them in the paper.

Results

To be discussed during Measuring Behavior 2016.

We hypothesize that the Combined Baby Model will outperform all other models on the two test sets. We also hypothesize that the Google Images Baby Model will perform the best on its test set equivalent. Finally, we hypothesize that the the Experimental Baby Model will perform best on its test set equivalent.

Future Directions

To be discussed during Measuring Behavior 2016.

Proposed future direction is to account better for movement in a video, i.e. create a more robust baby face combining date collected from 3D as well as 2D inputs [17]

Reference

1. Moeslund, T. B., & Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer vision and image understanding* **81**(3): 231-268.
2. Cheng, D., & Vertegaal, R. (2004). An eye for an eye: a performance evaluation comparison of the LC technologies and Tobii eye trackers. *Eye Tracking Research & Application: Proceedings of the 2004 symposium on Eye tracking research & applications* **22**(24): 61.
3. Valstar, M. F., Jiang, B., Mehu, M., Pantic, M., & Scherer, K. (2011). The first facial expression recognition and analysis challenge. *International Conference on Automatic Face & Gesture Recognition and Workshops*, 921-926.
4. Landis, C. (1924). Studies of emotional reactions: II. General behavior and facial expression. *Journal of Comparative Psychology* **4**: 447-509
5. Peters, J. W., Koot, H. M., Grunau, R. E., de Boer, J., van Druenen, M. J., Tibboel, D., & Duivenvoorden, H. J. (2003). Neonatal Facial Coding System for assessing postoperative pain in infants: item reduction is valid and feasible. *The Clinical journal of pain* **19**(6): 353-363.
6. Kohut, S. A., Riddell, R. P., Flora, D. B., & Oster, H. (2012). A longitudinal analysis of the development of infant facial expressions in response to acute pain: Immediate and regulatory expressions. *Pain* **153**(12): 2458-2465.
7. Hetherington, M. M., Madrelle, J., Nekitsing, C., Barends, C., de Graaf, C., Morgan, S., & Weenen, H. (2016). Developing a novel tool to assess liking and wanting in infants at the time of complementary feeding–The Feeding Infants: Behaviour and Facial Expression Coding System (FIBFECS). *Food Quality and Preference* **48**: 238-250.
8. Messinger, D. S., Duvivier, L. L., Warren, Z. E., Mahoor, M., Baker, J., Warlaumont, A., & Ruvolo, P. (2014) Affective Computing, Emotional Development, and Autism. *Handbook of Affective Computing* (NY), 516-536
9. Oster, H. (2005). The repertoire of infant facial expressions: An ontogenetic perspective. *Emotional development*, 261-292.
10. Oster, H. (2015). Baby FACS: Facial Action Coding System for infants and young children. *Unpublished monograph and coding manual*. New York University.
11. Bagnato, L., Sorci, M., Antonini, G., Baruffa, G., Maier, A., Leathwood, P., & Thiran, J. P. (2007). Robust infants face tracking using active appearance models: a mixed-state CONDENSATION approach. *Advances in Visual* Computing (Springer Berlin Heidelberg), 13-23.
12. Fang, C. Y., Huang, L. J., & Chen, S. W. (2012). An Infant Facial Expression Recognition System. *Proceedings of the 25th IPPR Conference on CVGIP* (Sun Moon Lake, Taiwan)

13. Pal, P., Iyer, A. N., & Yantorno, R. E. (2006). Emotion detection from infant facial expressions and cries. In Acoustics, Speech and Signal Processing, *ICASSP 2006 Proceedings* **2:** II.
14. den Uyl, M., & van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. In L. P. J. J. Noldus, F. Grieco, L. W. S. Loijens, & P. H. Zimmerman (Eds.), *Proceedings of Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research* (Wageningen, The Netherlands: Noldus Information Technology), 589–590.
15. Cootes, T. F., & Taylor, C. J. (2004). Statistical models of appearance for computer vision. *Imaging Science and Biomedical Engineering* (University of Manchester, Manchester, UK).
16. Den Uyl, T.M. (2008). A children's model for the FaceReader. *Bachelor's Thesis Report* (VicarVision, Amsterdam, the Netherlands and Utrecht University, Utrecht, the Netherlands).
17. Asthana, A., Saragih, J., Wagner, M., & Goecke, R. (2009). Evaluating aam fitting methods for facial expression recognition. 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII 2009), 1-8.
18. Jeni, L. A., Cohn, J. F., & Kanade, T. (2015). Dense 3D face alignment from 2d videos in real-time. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition* (Ljubljana, Slovenia).

Use of Noldus Observer XT for Measuring Environmental Impacts on Individuals with Autism Spectrum Disorders

Shireen M. Kanakri[1], Haitham M. Fawaz[2], and Megan D. Schott [3]

[1]Department of Family and Consumer Science, College of Applied Technology, Ball State University, Muncie, IN, USA, smkanakri@bsu.edu
[2]Department of Electrical Engineering, Dwight Look College of Engineering, Texas A&M University, College Station, TX, USA, haithamkanakri@tamu.edu
[3]Psychology Department, College of Applied Technology, Ball State University, Muncie, IN, USA, mdschott@bsu.edu

Abstract

The purpose of the Autism Design Lab research is to understand what design changes in classrooms and autism treatment settings will best facilitate effective learning for children with autism spectrum disorder. In order to empirically identify what aspects of the classroom environment are best suited for students with ASD, the current classroom environment must be measured in a controlled way. The Autism Design Lab uses Noldus Observer XT software and Biopac system to do so, specifically for children with autism. Researchers collected data on seven self-stimulatory behaviors from high-functioning children with autism at two autism treatment centers. Behaviors recorded were repetitive motor movements, repetitive speech, ear covering, hitting, loud vocalizations, blinking, and verbally complaining in relation to decibel levels. The most powerful thing about this method is that this lab allowed researchers to observe and record all behaviors while simultaneously measuring loudness with the dB meter, and in the future will allow researchers to record variations in light intensity. This allows researchers to measure the effect of acoustics and lighting on the behavior of children with ASD. Future research will use a movable room unit consisting of acoustical panels, LED lighting, speakers with specific sounds, and a camera in addition to the Noldus Obser XT software and Biopac system to manipulate lighting and noise.

Introduction

The Center for Disease Control and Prevention estimates that one in 68 eight-year-old children are diagnosed with autism spectrum disorder (ASD) [1]. ASD is a neurodevelopmental disorder involving characteristics such as delayed verbal and social skills and self-stimulatory behaviors, which are repeptitive behaviors that serve no real purpose. Children with ASD also often experience auditory hypersensitivity, known as hyperacrusis [2]. Due to this oversensitivity to noise, children with ASD often cover their ears when noise levels rise. Autistic children also tend to have problems with modulating sensory input and shifting attention between stimuli [3]. While it is clear that noise has a negative impact on the learning capcity of individuals with ASD, little has been done to address challenges related to environmental distractions in classroom settings [4]. In order to generate classroom design guidelines to support the needs of autistic children, we collected data from several special needs classrooms and measured how acoustics impacted the behavior of children with autism.

Survey Design

The first phase of our research began in 2011 at Texas A&M University. 75 teachers from three schools for children with moderate to high functioning ASD in Houston, Texas completed a survey that addressed the impact of architectural design elements on autistic behavior in a classroom setting. Items for the survey were generated based on input from multiple sources, including interviews with parents, interviews with teachers, and a literature review that focused on anecdotal reports from people with autism such as Daniel Tammet and Temple Gradin. The following topics were addressed in the survey:

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

1. In addition to air conditioner sounds, echoes, sounds from children in the classroom, sounds from other classrooms, and traffic noise, are there other acoustical environmental conditions that you feel negatively affect children with autism?

2. Please rank the following regarding how negatively they impact the children's behavior: air conditioner, echoes, sounds from children in the classroom, sounds from other classrooms and traffic noise.

3. Are there aspects of the physical environment that you believe reduce the noise levels?

4. Describe the positive and negative acoustical qualities of the following types of rooms: classroom, common area, art room, computer lab and library, music and drama rooms, PE room.

5. As a teacher for children with autism, please evaluate the importance of carpet, wood panels, wood chairs, soundproofing, these aspects of the learning environment.

6. What types of behaviors do you see children doing that indicate that they are impacted by noise?

7. Do the children in the class ever attempt to reduce the noise by covering their ears or using 'ear defenders'?

Based on the results from the survey and extensive literature review, we identified five behaviors typically exhibited by children with ASD: repetitive movement (stereotypy), hitting in response to sound, producing loud sounds, blinking eyes, and complaining. We also identified several potential design considerations that may be used to manage noise, including the use of thick walls, carpet, wood chairs and tables, and wood panels. The second phase of our research focused on measuring the behaviors of children with ASD in relation to noise levels, while the next phase of our research will focus on the manipulation of some of these design features and measuring how they influence behavior.

Measuring Autistic Behaviors

Data collection for the next phase of our research began in 2012 and focused on measuring the behaviors of children with ASD in relation to dB levels in a classroom setting. Participants were second and third grade children with high levels of autism in four different classrooms from two schools in Texas (see Figure 1). The purpose of the study was to measure how noise levels influence the behaviors identified in the survey and literature review.

Figure 1. School 1, classrooms 1 and 2.

A video recorder linked to a decibel meter was located in the classroom during the behavioral observation time. The researcher defined the range of the noise level which would be observed from the pilot study results. The video camera recorder was a Logitech - C920 Pro Webcam – Black. This type of camera had different systems and still image resolution up to 5.0 megapixels for intense clarity. Video captured up to 720p for lifelike detail and motion.

The video camera recording system had a two-rotary head, helical scanning system. The audio recording system had rotary heads and a PCM system. The recording time was up to 6 hours daily (DVM60ME). The lens dimensions were 10x (optical), f=1/4 to 25/16 inches, 1 5/8 to 169/16 inches when converted into a 35-mm still camera. It also had input and output connectors, including S video output, Audio output. The decibel meter model was an "1800 Precision Integrating Sound Level Meter" and it contains an octave band filter model OB-100 and 1/1 – 1/3 octave band filter model OB-300. This meter functions as a precision sound level meter, impulse or integrating sound level meter. In all modes, the model 1800 delivers type 1 accuracy include laboratory, industrial, community and audiometric measurement and analysis [5]. To view materials used in this study, see Figure 2.

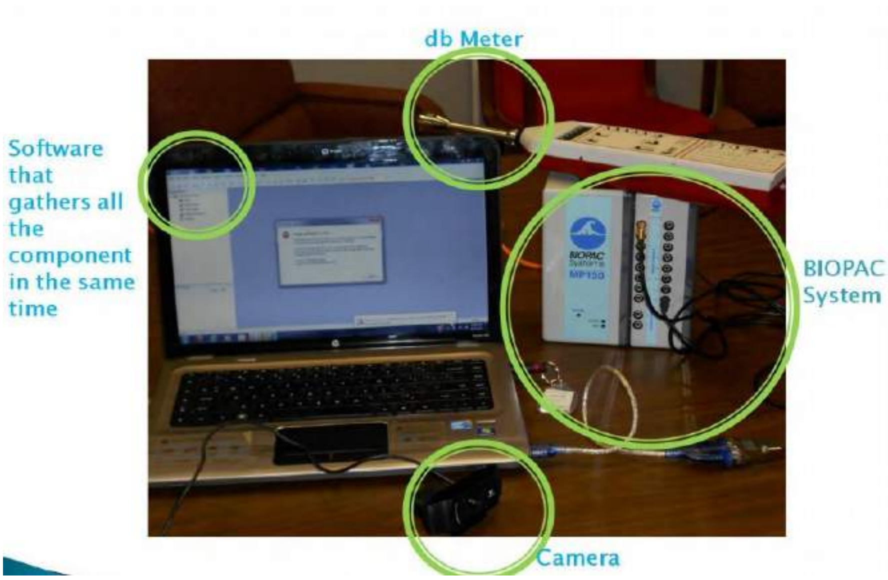


Figure 2. Noldus Observer XT Software, Logitech - C920 Pro Webcam – Black, 1800 Precision Integrating Sound Level Meter, and Biopac system.

A 12-hour pilot study was conducted to test the tools and the five hypothesized behaviors to be observed based on an extensive literature reviwew (repetitive movement (stereotypy), hitting in response to sound, producing loud sounds, blinking eyes, and complaining). After the 12 hours of observation in the pilot study, two additional behaviors were deemed important to track: covering ears and repetitive speech. In total, all seven behaviors were tracked in this study. After integrating the information gathered from the pilot study, observations ran from 8:30am until 2:30pm for 48 days. Each week, approximately 20 hours of observations were recorded. In order to obtain 64 hours of data from each classroom, 6 weeks of observations took place, with a two day break between each observation period. In total, in addition to the half-day pilot study, the observation period was seven weeks.

The observation utilized the three techniques mentioned above. In the beginning the researcher used the Noldus Observer XT software to record the number of times the children exhibited the identified seven behaviors. A measurement was done every 30 seconds. Every time the children exhibited one of the behaviors it was recorded by the researcher as a colored mark indicating the behavior and the child's gender.

Every ten seconds a video recorder that was linked to a decibel meter recorded all the activities and the sound levels inside each classroom during the observation period. At the end of each day of observation a short questionnaire was submitted to the teachers of both classes of the study in which they were asked to give their opinion and ideas about the classroom environment.

The researcher also recorded the source, volume and duration of the noise every 10 seconds in order to associate the noise types with the times of behavior. A measurement was taken every other 10 seconds. In this way the researcher would have a chance to record and write any notes about the behaviors and the noise properties of that 10 second interval. Every time a child responded with one of the behaviors it was recorded by the researcher as a child code letter to indicate the behavior times. The researcher would use the behavioral observation sheets from observer XT to collect the data.

Every 10 seconds a video recorder that was linked to one decibel meter recorded all activities and sound levels inside each classroom during the observation time. Every 5 minutes the researcher checked the video camera and decibel meter, and then the researcher made notes to match them with the equipment readings.

The researcher used software that linked the behaviors with the decibel meter readings, so in this way the behavioral record was tied to decibel reading for the noise. Behavioral recording and sound recordings were analyzed offline using Noldus Observer XT. At the end of each day of observation a short questionnaire was given to the teachers of both classes of the study allowing them to express their opinion and ideas about the classroom environment. To view the behavior recordings analysis using the Noldus XT Observer system, see Figure 3.

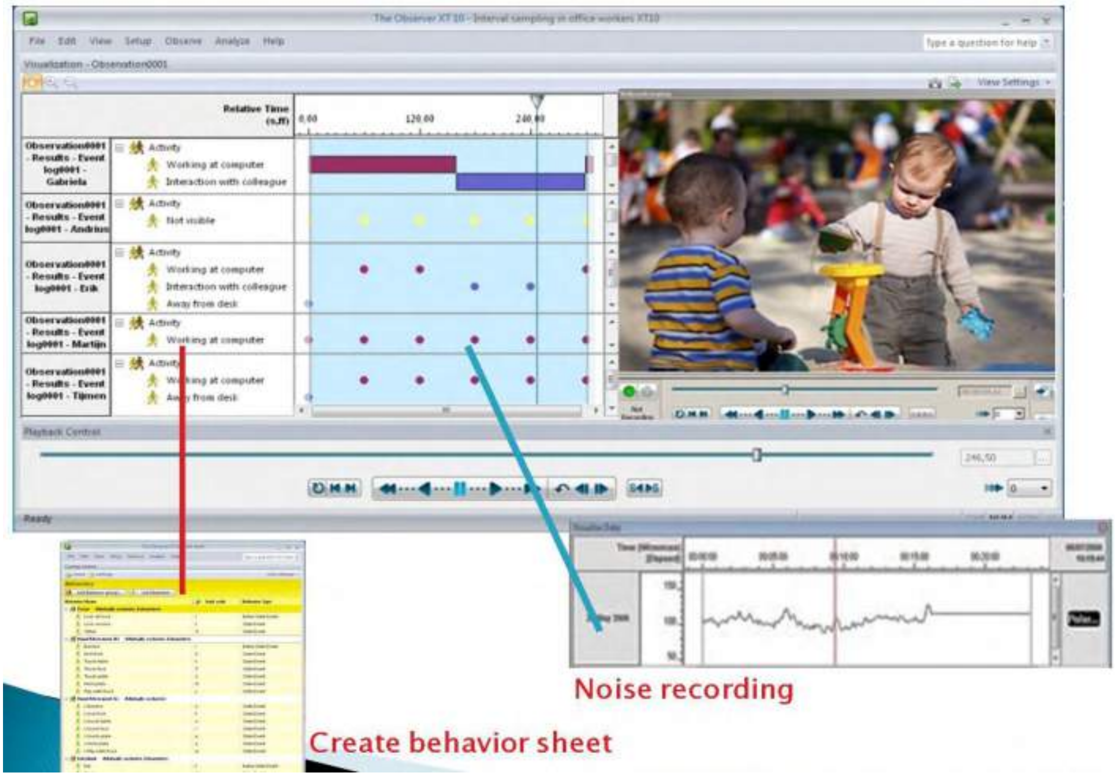


Figure 3. Behavior recording analysis using the Noldus Observer XT system.

Data Analysis

Following data collection, the data was cleaned and entered into a Microsoft Excel sheet and then inputed into a Statistical Package for the Social Sciences data file to be analyzed. Because the Noldus XT and Biopac systems allowed behaviors and dB levels to be recorded simultaneously, we were able to do several statistical analyses.

We were able to analyze the frequency of behaviors by decibel level, which allowed us to break the data down into quiet (41-55 dB), loud (55-70 dB), and poetentially damaging (> 70 dB) sound levels. Pearson correlation coefficients were also computed to examine the relationship between dB level and each of the seven behaviors.

Results

In general, results show a significant positive correlation between decibel levels and most behaviors observed in this study. When data across decibel levels and classrooms for Schools I and II are aggregated, instances of complaining, repetitive speech, hitting, producing loud sounds, repetitive motor movements, and covering ears show a tendency to increase as noise levels rise. Blinking eyes was the only behavior that did not follow this general pattern. While data indicates that a general positive relationship exists between decibel level and behavioral occurrences, disaggregation of the data illustrates the variability that exists with regard to specific decibel ranges. Repetitive motor movements occurred much more frequently in the Loud Range (55-70 dB), which we hypothesize may indicate a sign of auditory distress for the students. In addition, repetitive speech, producing loud sounds, covering ears, blinking eyes, complaining, and hitting others also occurred with greater frequency in this range, which may also be a sign of auditory distress. For the Potentially Damaging Range (>70), repetitive speech and covering ears occurred much more frequently than in the previous ranges, while repetitive motor movements decreased. If the function of the students' behavior in this range is to attempt to escape from the overwhelming auditory input, it is possible that increased repetitive speech may be a more salient way for them to express their distress or manage the discomfort than engaging in repetitive motor movements. A functional behavior assessment would be needed to validate such hypotheses.

Discussion

The Noldus Observer XT research software is commonly used in autism research, and Noldus even has references for 14 scientific articles published in journals listed on their website [6]. However, what is novel about our approach to measuring behavior is that by connecting the Noldus Observer XT software to the Biopac system, we are able to measure the behaviors and dB levels at the same time. This allows us to draw conclusions about how noise and acoustics impact the autistic user's behavior.

Future Research

Up to this point, our research has focused on measuring seven autistic, self-stimulatory behaviors in relation to dB levels in classroom settings. Moving forward, our Autism Design Lab will use Noldus Observer XT software and a moveable room unit consisting of acoustical panels, LED lighting, Biopac system, speakers with specific sounds, and a camera. This moveable room unit with controlled lighting and noise will further current research by inviting children with autism and their parents to be observed under various environmental conditions manipulated and controlled by this unit at Ball State University. The most powerful thing about this method of research is that all behaviors, dB meter variatoins, and light intensities will be recorded simultaneously so we can observe how these variables affect the behavior of autistic children.

References

1. Center for Disease Control and Prevention. (2014). Prevalence of autism spectrum disorder among children aged 8 years (MMWR Publication No. 63(SS02)). Atlanta, GA: Center for Disease Control and Prevention.
2. Baguley, D., Andersson, G., McFerran, D., & McKenna, L. (2013). Tinnitus and Hyperacusis in Childhood and Adolescence. *Tinnitus: A Multidisciplinary Approach, Second Edition*, 208-214.
3. Courchesne, E., Lincoln, A. J., Yeung-Courchesne, R., Elmasian, R., & Grillon, C. (1989). Pathophysiologic findings in nonretarded autism and receptive developmental language disorder. *Journal of Autism and Developmental Disorders*, **19**, 1-17.

4. Mostafa, M. (2008). An architecture for autism: Concepts of design intervention for the autistic user. *International Journal of Architectural Research*, **2**, 189-211.
5. Quest. (1999). Instructions for Model 1800 Precision Integrating Sound Level Meter and Model OB-100 Octave Band Filter and Model OB-300 1/1 – 1/3 Octave Band Filter. <http://multimedia.3m.com/mws/media/760875O/1800-sound-level-meter-user-manual.pdf> Accessed 22 January 2016.
6. Noldus Information Technology. Autism research. <http://www.noldus.com/human-behavior-research/application-areas/autism-research>. Accessed 21 January 2016.

Longitudinal Video-based Measurement and Coding of Leader/Follower Behavior and Team Performance

D. H. van Dun[1,2] and C. P. M. Wilderom[1]

[1] Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede, the Netherlands
[2] House of Performance, Utrecht, the Netherlands
d.vandun@hofp.nl

Purpose

Effective leaders are known to display exemplary behavior at work, and thereby have a strong effect both on the behaviors of their followers and their teams' performance [1, 2]. Although this belief is widespread among a vast majority of leadership scholars and management practitioners, we note that this idea is predominantly drawn from perceptual data, including researchers' field notes. In order to advance our knowledge of the effects of effective leaders' behavior, more nuanced and objective video-observations of leader and follower behaviors are called for [3, 4]. Such studies are also advised to include objective measures of team performance, in order to get a better understanding of the effects of effective leader and follower behavior on team performance [5, 6]. The purpose of the present field study is to examine —over time— the effects of effective team leader behaviors on both the behaviors of their followers and on team performance, because work teams may fluctuate in terms of their performance over time.

Methodology

In order to select high-performing teams, we sent out a call for their (self-)nomination through a diverse set of known Dutch management media and networks. For each nominated team, we retrieved, from their archives, objective team performance data and held interviews with key informants. After getting their followers' consent for the video part of the study, all team leaders agreed to participate in the study. All five selected teams worked in different organizations: we studied two manufacturing teams (the "Truck" and "Commodity" team) and three in the service sector (the "Insurance" and "Mail" team, as well as one team operating in the public sector: the "Government" team). A moderate level of attrition appeared among the total of 63 participants over time (i.e., 5 team leaders and their followers).

At two points in time (year 1 and year 3), each of the teams was visited by a single researcher, during one full work week. The three field researchers in our study were carefully selected Master and Ph.D. students in Psychology or Business Administration. They were trained extensively before the study in the field protocol and (video) observation techniques [7, 8]. Moreover, the researchers wore company clothing or dressed unobtrusively. It is important to note here that the focal team leaders and their followers were accustomed, to some extent, to being monitored by others: Each team actively used visual management tools and daily start-of-shift meetings in order to inform followers of the up-to-date team performance. In three of the five teams, this information was continually updated and displayed on computer screens in a central area in the workplace.

We took two full days out of each observation week to get accustomed to the leader and his/her team through "observant participation" [9]. On the third day we introduced the video camera in the workplace and recorded all that happened in the workplace, thus enabling the team leaders and their followers to get used to the camera. Followers noted that they quickly forgot that they were being recorded. The team leaders also reported that neither they, nor their followers, behaved differently due to our presence. Our actual data collection started on day 4 and 5; we video-taped two different work settings:

8. We recorded individual team leaders or their followers *while working behind their desks or workbench,* including their naturally occurring interactions with their colleagues, over the course of one hour (twice per person) [3, 7, 8, 10], even at the coffee machine: with a method named "video-shadowing" [11].

9. *Regular team meetings* that included both the daily start-of-shift meetings (about 15 minutes) and the weekly monitoring meetings (about 60 minutes). In both instances we taped both the facial and verbal expressions of the team leaders and their followers. We used two video cameras: one on a tripod, pointed at the leader, and a hand-held one that captured the followers' behavior.

We collected a total of 62.5 hours of video data and three years of team's performance measured from archival data, based on their own key performance indicators.

Data Analysis

We deliberately chose to use video to capture leader and follower behaviors, because we were able to watch the videos multiple times, which enabled us to analyze the behaviors at a very detailed level-of-analysis [12, 13]. The video data was examined with The Observer XT software [14]. Following a detailed codebook that was slightly adapted from Hoogeboom and Wilderom [15], two independent raters scored the minute behaviors of both the team leaders and their followers that appeared in the videotaped episodes (the technical details of this procedure will be explained during our conference presentation). The codebook consisted of 14 mutually-exclusive behaviors that involve the following three spectra:

10. Task-orientated behavior (i.e., correcting, delegating, task-performance monitoring, informing, visioning, and structuring the meeting).

11. Relations-oriented behaviors (i.e., agreeing, providing individual consideration, intellectually stimulating, and active listening).

12. Counter-productive behaviors (i.e., showing disinterest, defending one's own position, providing negative feedback, and disagreeing).

After coding each tape, the two raters compared the parts that deviated from each other when they were automatically flagged by The Observer XT software. Finally, inter-rater reliabilities were retrieved: 97.9% at year 1 and 100% at year 3.

We calculated each team's average, standardized behavioral frequencies at *two* points in time for *four* different work situations: team leader behavior during meetings, follower behavior during meetings, team leader behavior during daily work, and follower behavior during daily work. A series of repeated measures analyses explored the differences in behavioral frequencies measured in year 1 and 3 in relation to team performance in each of the five teams.

Results and Conclusions

The mixed-methods approach, including both the quantification of team leaders' and their followers' minute behavior at work, as well as the retrieval of key team performance indicators over three years, illuminated a leader-behavioral cascade effect. We illustrate this effect in Table 1, which displays the team-specific average behavioral frequencies, both for its team leader and followers, and divided per video-observed work situation (meetings or daily work). From Table 1 it can be seen that the Truck and Commodity team leaders display relatively often active listening and information sharing, and *little* task-performance monitoring and counterproductive behaviors towards their followers. Table 1 also shows that, over time, their followers actually take over their task-monitoring behavior, especially during meetings. In effect, these teams demonstrate higher team performance over time, e.g. higher productivity and lower defect rates. On the other hand, the Insurance and Mail team leaders who display somewhat *more* task-performance monitoring at work, face followers who display less task-monitoring behavior during meetings; these teams do not sustain their initial high performance. In other words, if team leaders facilitate employees, through information and listening, their employees feel safe to monitor their own performance and engage in efforts to further improve their team performance.

Although the video data was rather time-consuming to analyze, we were able to build upon the more exact measurement of leader and follower behavior through the use of video: beyond perceptual (survey) measures. Our video-behavioral approach is not only likely to contribute to more accurate management and leadership theory, but will also inform, in practice, those who are actually in charge of a work team and who want to become more effective leaders. The findings may help managers receive better or more objective feedback-based guiding and training in terms of their display of specific behaviors at work. Viable next research steps include video-based behavioral pattern analyses, or video-analyses of leader-follower turn-taking rhythms among the leaders and followers in both effective and ineffective work teams.

References

1. Zaccaro, S.J., Rittman, A.L., Marks, M.A. (2001). Team leadership. *The Leadership Quarterly* **12**, 451-483.
2. Mathieu, J., et al. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management* **34**, 410-476.
3. Yukl, G. (2012). Effective leadership behavior: What we know and what questions need more attention. *Academy of Management Perspectives* **26**, 66-85.
4. Pentland, A. (2012). The new science of building great teams. *Harvard Business Review* **90**, 60-70.
5. Hackman, J.R. (1987). The design of work teams. In: Lorsch, J. (Ed.) *Handbook of Organizational Behavior*. Prentice-Hall: Englewood Cliffs, NJ, 315-342.
6. Marks, M.A., Mathieu, J.E., Zaccaro, S.J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review* **26**, 356-376.
7. McDonald, S. (2005). Studying actions in context: A qualitative shadowing method for organizational research. *Qualitative Research* **5**, 455-473.
8. Vásquez, C., Brummans, B.H.J.M., Groleau, C. (2012). Notes from the field of organizational shadowing as framing. *Qualitative Research in Organizations and Management: An International Journal* **7**, 144-165.
9. Czarniawska, B. (2008). Organizing: How to study it and how to write about it. *Qualitative Research in Organizations and Management: An International Journal* **3**, 4-20.
10. LeBaron, C.D. (2008). *Microethnography.* In *The International Encyclopedia of Communication*, Donsbach, W., Editor. Blackwell Publishing: Blackwell Reference Online.
11. Czarniawska, B. (2007). *Shadowing, and other techniques for doing fieldwork in modern societies*. Malmö, Sweden: Liber AB.
12. Liu, F., Maitlis, S. (2014). Emotional dynamics and strategizing processes: A study of strategic conversations in top team meetings. *Journal of Management Studies* **51**, 202-234.
13. Paroutis, S., Franco, L.A., Papadopoulos, T. (2015). Visual interactions with strategy tools: Producing strategic knowledge in workshops. *British Journal of Management* **26**, S48-S66.
14. Noldus, L.P.J.J., et al. (2000). The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments, & Computers* **32**, 197-206.
15. Hoogeboom, A.M.G., Wilderom, C.P.M. (2015). Effective leader behaviors in regularly held staff meetings: Surveyed vs. videotaped and video-coded observations. In: Allen, J., Lehmann-Willenbrock, N., Rogelberg, S.G. (Eds.), *The Cambridge handbook of meeting science*. Cambridge University Press: Cambridge, United Kingdom, 381-412.

Table 1. Mean Frequencies (in %) of Video-Observed Behaviors During Meetings (M) and Daily Work (DW), Year 1 (T1) versus Year 3 (T3)

2.	Truck								Commodity							
	Team Leader's Behavior				**Followers' Behavior**				**Team Leader's Behavior**				**Followers' Behavior**			
	M T1	**M T3**	**DW T1**	**DW T3**	**M T1**	**M T3**	**DW T1**	**DW T3**	**M T1**	**M T3**	**DW T1**	**DW T3**	**M T1**	**M T3**	**DW T1**	**DW T3**
1. DIS	0	0.93	0	0.92	27.94	21.59	0	0.94	0	0	0	1.47	0	11.31	0	1.46
2. TPM	8.87	6.70	10.21	5.97	14.38	9.55	7.40	3.34	5.92	0	13.00	0.78	11.11	33.93	10.00	1.22
3. Informing	23.22	33.90	16.80	19.12	24.67	10.80	10.16	10.13	23.67	42.46	18.90	32.22	31.48	39.29	8.33	11.79
4. Visioning	18.28	1.92	9.57	0	1.47	0	0.55	0	5.33	0	10.59	0.23	26.54	0	0	0
5. Agreeing	1.61	5.56	4.55	4.11	1.47	9.90	0.56	2.04	0.59	9.01	0.95	5.22	4.94	11.31	0	11.57
6. PIC	6.35	9.54	6.35	10.88	28.60	37.72	28.48	25.69	9.47	1.56	3.72	4.75	9.88	4.17	13.83	12.06
7. AL[a]	20.48	35.76	33.77	34.57	-	-	13.82	26.72	42.01	44.03	34.05	39.00	-	-	28.20	26.77
	Government								**Insurance**							
	Team Leader's Behavior				**Followers' Behavior**				**Team Leaders' Behavior**				**Followers' Behavior**			
	M T1	**M T3**	**DW T1**	**DW T3**	**M T1**	**M T3**	**DW T1**	**DW T3**	**M T1**	**M T3**	**DW T1**	**DW T3**	**M T1**	**M T3**	**DW T1**	**DW T3**
1. DIS	0	0.38	0	1.24	3.26	20.95	0	0.74	0.24	0.58	0.23	0	0.43	19.39	0	0.40
2. TPM	13.00	18.76	10.33	5.33	12.06	11.35	10.44	7.29	19.14	13.75	9.18	10.48	8.03	5.62	2.91	0.53
3. Informing	19.40	19.52	22.28	11.73	30.51	23.52	12.73	10.44	10.25	21.93	9.41	20.49	37.37	39.26	12.22	23.31
4. Visioning	4.30	0.38	5.83	0	14.48	0	5.97	0	11.86	1.08	6.53	1.02	25.07	0	11.12	0
5. Agreeing	1.26	5.17	2.47	9.77	5.27	19.26	0.30	5.84	1.73	15.74	2.33	7.40	7.03	16.61	0.85	6.97
6. PIC	9.67	10.18	11.15	23.08	26.91	15.14	13.50	20.68	8.86	4.14	17.43	10.51	12.94	5.29	18.71	10.55
7. AL[a]	38.59	36.79	39.67	38.43	-	-	34.80	27.88	34.37	36.08	34.98	40.89	-	-	33.19	35.15
	Mail															
	Team Leader's Behavior				**Followers' Behavior**											
	M T1	**M T3**	**DW T1**	**DW T3**	**M T1**	**M T3**	**DW T1**	**DW T3**								
1. DIS	0	0	0.13	1.46	33.12	70.48	0	1.51								
2. TPM	6.44	10.73	10.20	7.49	11.84	0.29	7.94	1.57								
3. Informing	31.43	31.56	19.65	19.49	19.58	8.74	11.01	4.89								
4. Visioning	13.61	0	9.45	0.23	5.06	0	4.92	0								
5. Agreeing	2.98	1.19	2.12	5.73	2.46	8.36	1.00	4.11								
6. PIC	5.10	24.48	6.42	10.46	17.13	11.26	19.96	20.32								
7. AL[a]	16.48	5.90	35.30	30.22	-	-	26.48	33.81								

Note. DIS = Showing disinterest; TPM = Task-performance monitoring; PIC = Providing individual consideration; AL = Active listening. The coding scheme consisted of a set of 14 mutually exclusive behaviors with which we reliably coded all video data; only the behaviors with relevant time differences are listed in this table. Each mean behavioral frequency in this Table is based upon n = 2 or higher (except for the Commodity team which had only one meeting at T1).

[a] Because the videotape had not always captured followers' precise facial expressions during the meetings, we could not reliably code followers' "active listening" during those filmed episodes. Hence, we left this behavior out of the analysis of the meetings.

Pupillary Responses of Asian Observers in Discriminating Real from Fake Smiles: a Preliminary Study

M.Z. Hossain[1], T.D. Gedeon[1], R. Sankaranarayana[1], D. Apthorp[2], A. Dawel[2]

[1]Research School of Computer Science, Australian National University, Canberra, Australia
[2]Research School of Psychology, Australian National University, Canberra, Australia

Abstract

This work is a part of a research effort on detecting the differences between human responses to real and fake smiles, by watching observers of different smiling video stimuli (displayer's smiles). Pupil diameter was recorded from 10 Asian observers (6 males and 4 females), while watching 9 real smile stimuli and 10 fake smile stimuli. The preliminary analysis on pupil data revealed that the pupil size increased more for fake smile stimuli as compared to real smile stimuli in the case of male observers, and vice versa in the case of female observers. To the best of our knowledge, most of the work on discriminating real from fake smiles was analyzed from the smile displayer's point of view. Our results were found from observers' pupillary responses. If we consider fake smiles as negative emotions and real smiles as positive emotions, then this outcome may be comparable with previous findings on pupil dilation. Our results show that analyzing an observer's pupillary responses may help to understand the displayer's actual state of mind.

KeyWords: Real Smile, Fake Smile, Pupil Size Variation, Asian Observers, Video Stimuli

Introduction

Human smiling would appear to be one of the most complex, but seemingly straightforward facial displays. Although smiles reflect positive affect [1], this can be from a variety of emotions [2]. People do not smile only for happiness, but also when they are socially anxious, embarrassed, depressed, surprised and even more [3]. In many cultures, cheerfulness is mandatory and workers are required to smile as part of their jobs. Furthermore, it has been noted that smiling is less a sign of original emotion than a social display meant for others [4]. In this paper, we refer to posed, acted and non-happy smiles as fake smiles and only consider smiles grown from happiness as real smiles. In this context, understanding the meaning of a smile can help us to know a displayer's actual state of mind.

In the past, Valstar et. al. [5] distinguished real from fake smiles by analyzing 202 smiling videos directly. The smiling characteristics of virtual agents were studied in [6]. The displayers' perceived meanings regarding smiles were found to be related to specific characteristics [7]. The dynamics and morphological characteristics of displayers' smiles were analyzed to disambiguate real from fake smiles in face to face interaction [3]. They addressed displayers' smiling characteristics to discriminate real from fake smiles. In our work we try to analyze an observer's pupillary responses while they watch a (video of a) displayer's smile.

The human pupil may dilate for various reasons, including memory load, stress, pain, different emotional stimuli and so on [8]. Chapman et al. [9] reported that pupil dilation changes significantly with the increase of pain intensity. Kang et al. [10] observed that pupil dilation was lager when people were more curious about the answer. Wang et al. [11] found that pupil dilation was proportional to the 'size' of a lie. The pupil size variation also reflects the judgments of males on females and vice versa. For example, the pupil size was increased when observers were shown facial stimuli of opposite sex, and larger pupils of females evoked more positive feelings in males [12]. Thus measuring pupillary responses would offer a good method for smile detection, because no sensors need to be attached to the observer, and certainly not to the displayer.

The above mentioned findings on pupillary responses suggest the measurement of pupil size may be useful to discriminate real from fake smiles. Our study using observers' pupillary responses actually reflect some of how and what a person is thinking about another's smiling behavior. This preliminary study offered us a good outcome in discriminating real from fake smiles with an indication of gender differences. We found evidence that Asian males differ from Asian females according to their pupillary judgments.

Materials and Methods

Observers

Ten healthy, Asian-background right-handed graduate students were participants as observers of video stimuli, with a mean age of 29.8 ± 4.8 for males, and 34.7 ± 4.8 for females ($\text{mean} \pm \text{SD}$). Each observer had normal or corrected to normal vision, and provided their written consent prior to his/her voluntarily participation. An approval from the Australian National University's Human Research Ethics Committee was granted for this study.

Stimuli

It is well known that video is more effective and easier than static images for human recognition purposes [17]. For that reason, twenty-five video stimuli were collected from three benchmark databases: ten from AFEW [13], five from MMI [14] and, ten from MAHNOB (five from HCI and five from Laughter) [15-16]. The collected video stimuli were processed using MATLAB R2014b to convert to grey scale, mp4 format and each lasting 10s from smile onset. This paper is based on the analysis of collected pupillary responses, while observing 10 AFEW and 9 MAHNOB video stimuli. Luminance of each stimulus over the frames in the videos was computed using the MATLAB SHINE toolbox [18]. The ten stimuli ($\text{mean luminance} = 70.2$ and $Standard\ deviation = 4.6$) of AFEW database were classified as fake smiles, because these were acted smiles by professional actors. On the other hand, participants' positive emotions (happiness) were elicited in the case of MAHNOB stimuli ($\text{mean luminance} = 75.8$ and $standard\ deviation = 2.1$) and were treated as real smiles. In the case of other 6 videos: we could find no clear specification or explanation in the source for one stimulus collected from the MAHNOB Laughter database so we could not classify it as fake or real; and the luminance of the MMI videos were much higher than the rest. For this reason, they were not considered for analysis purposes. The remaining videos had luminance which varied within 6 ALU (Arbitrary Linear Unit) for each stimulus. This consistent range allows us to confidently remove the effect of luminance from the pupil dilation.

Procedure

The observer was introduced to the laboratory and comfortably seated in a static chair in front of a 17.1 inch computer monitor. Their chair was moved forward or backwards to adjust the distance between the chair and eye tracker. The participant (observer) was told that their eye movements would be recorded during video stimuli presentation and instructed to limit their body movements to reduce unwanted artefacts in the signals.

The observer was instructed to track nine fixation points, which were displayed in the monitor for calibration purposes. The experiment was started with a short introduction of fake versus real smiles and finished with feedback through a web-based questionnaire. The pupil size variation of the observers was recorded from both eyes with a facelab (Seeing Machines) remote eye-tracker system. The sampling rate of the system is 60 Hz. Every stimulus presentation was followed by a five point (-2 to +2) Likert scale (see Figure 1) where the observer rated his/her experience as evoked by the stimuli using a computer mouse. The negative and positive ends of the scale represented fake and real smile stimuli respectively. The centre of the scale represented confusion. The ends and centre of the scales were labelled appropriately. Each viewing stimulus to the observer was rated by this scale and documented in a text file.

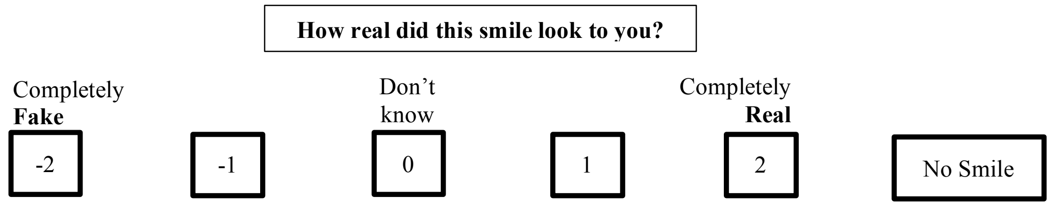


Figure 1. Five point Likert scale to accumulate the observer's self-report.

Data Processing

Samples where the pupil was obscured due to blinking were measured as zero by the eye tracking system, and cubic spline interpolation was applied to reconstruct the pupil size. Then, 10-point Hann moving window average was used to smooth the pupil signal. The minimum pupil diameter of each observer was subtracted from each pupil diameter of that observer, to reduce the effect of age among observers and to normalise for luminance in the videos which all had a consistent 6 ALU luminance range. The final pupil signal consisted of baseline corrected data averaged from both eyes over the whole 10s video stimulus interval.

All the data was analysed with paired sample two-tailed permutation test to report on significance levels. This test is performed to assess the statistical significance on waveform differences and produces p values for any number of time points [22]. Two-way analysis of variance (ANOVA) also introduced to check the effect of smile type, effect of sex, and their interactions.

Results

The timeline analysis of the pupil data revealed a common trend for each stimulus. Figure 2 illustrates the time point average of pupil diameter over observers when viewing all video stimuli. The pupil constricted from stimulus onset and reached a minimum within 1-2s, after which a sharp dilation started and continued till 3-4s. Then, either a smooth dilation or constriction started and continued, which is sustained in a consistent range, until the end of our analysis window. It is worth noting that the trends were separated according to real and fake smiling stimuli from about the 3s of stimuli onset. The paired sample permutation tests showed that pupil dilation differed significantly for real smile stimuli as compared to fake smile stimuli ($t = 4.56, df = 9, p = 0.051$), while mostly lower p values ($p < 0.1$) are found between 8.62s and 8.67s.

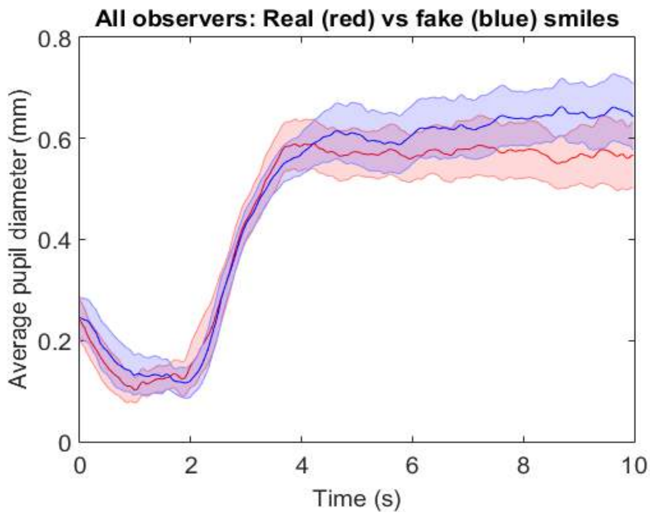
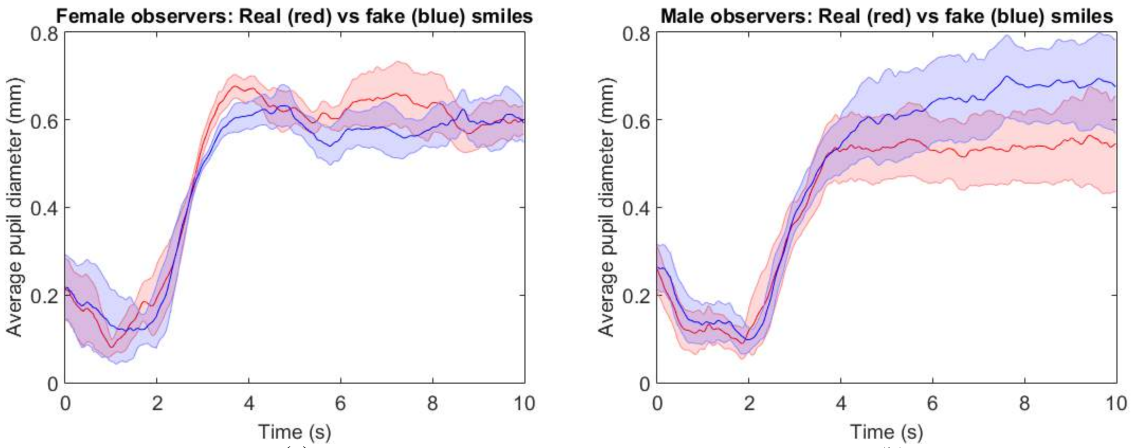


Figure 2. Average pupil diameter timelines for all stimuli over all observers.

The timeline of the pupil diameter changes was somewhat different for female and male observers (see Figures. 3(a) and 3(b)). The fake smile stimuli provoked the strongest pupil dilations for male observers, where the real smile stimuli provoked the strongest dilations for female observers. In the case of female observers, the difference of real from fake smiles started from 3s of stimulus onset and ended within 8s to 10s from stimulus onset (see Figure 3(a)). Paired Sample permutation test comparisons showed that the pupil diameter was not significantly smaller for fake smile stimuli than real smile stimuli ($t = 11.30, df = 3, p = 0.125$). In the case of male observers, the pupil dilates continuously from 4s of stimulus onset for fake smile stimuli, but no dilation or contraction for real smile stimuli (see Figure 3(b)). The pupil diameter was significantly larger for fake smile stimuli than real smile stimuli ($t = 6.41, df = 5, p = 0.063$), mostly lower p values ($p < 0.1$) are found from 7.75s to 7.93s and from 9.87s to 9.89s. The greater significant results may be found, if the number of observers is increased.



(a) (b)

Figure 3. Average pupil diameter timelines for all stimuli over (a) female and (b) male observers.

We have also calculated the average value of individual observers' pupil size, while watching fake and real smile stimuli. Individual analysis on the whole presenting stimulus interval verified that larger pupil dilations were detected by male observers in the case of fake smile stimuli as compared to real smile stimuli and vice versa in the case of female observers (see Figure 4). In this Figure, error bars indicate standard deviations.

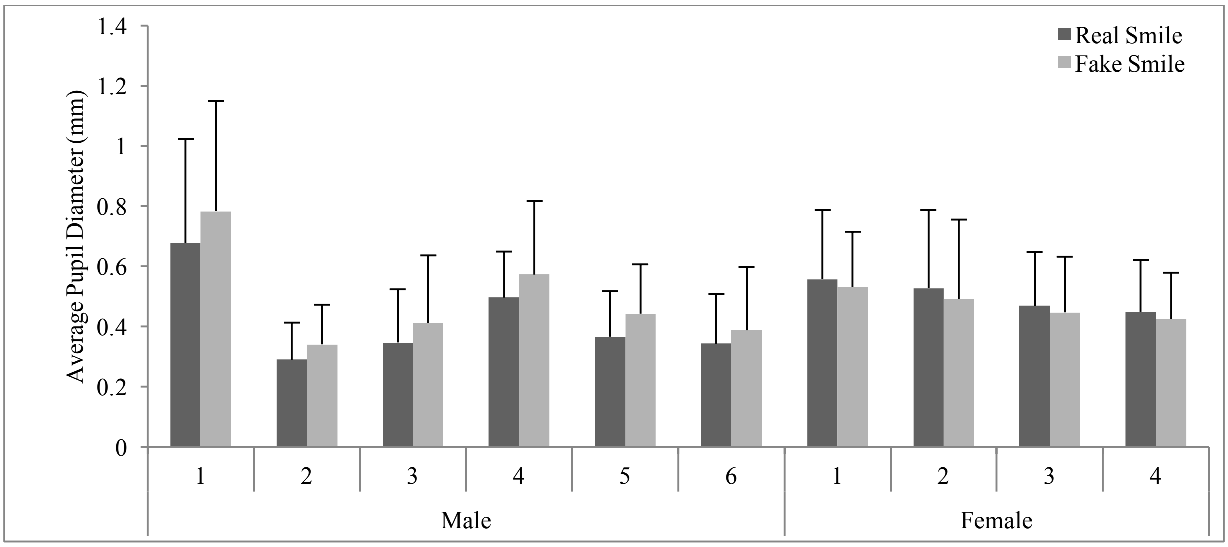


Figure 4. Average pupil diameter variation for individual observers.

Additional calculations on pupil size variation revealed that male observers have lower average values compared to female observers while watching real smile stimuli, and vice versa in the case of fake smile stimuli (see Figure 5). Here, error bars also denote standard deviations as in Figure 4.

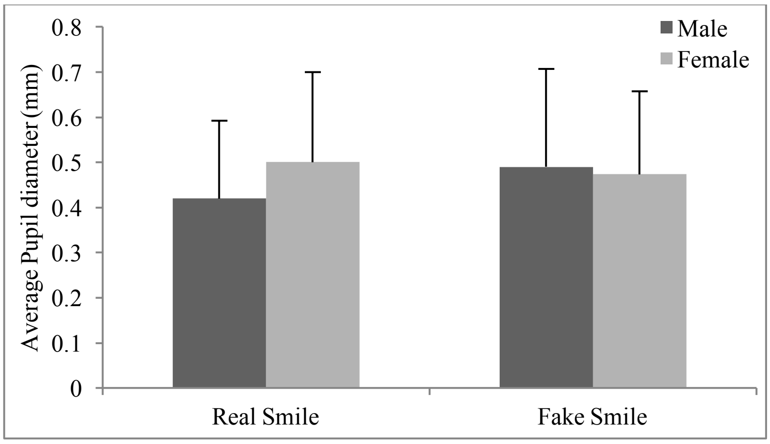


Figure 5. Average pupil diameter variation between female and male observers.

We also employed ANOVA test to check the effects of sex (male vs. female), smile type (real vs. fake), and their interactions on pupil data. The computed $F\ scores$ were $45.81, 20.42$ and 104.17 for the effects of sex, smile type and their interactions respectively, that have significant impact ($p < 0.001$ for each case) on pupil data. We have also observed the $R\ squared$ value that explained 79% variations in the pupil data by the combinations of these effects and their interactions.

Discussion

Our results showed larger pupil size dilation in response to fake smile stimuli in the case of Asian male observers as compared to the pupil dilation in response to real smile stimuli and vice versa in case of Asian female observers. These differences were found during smiling video stimuli observation. The analyses on individual observers' pupil dilation also revealed that real smile stimuli caused different pupillary responses as compared to fake smile stimuli.

The visual inspection of the timeline curves discovered that the individual observers' average pupil dilation over stimuli coherently followed a similar type of curve. A sudden decrease in pupil size from the stimulus onset was followed by a sharp increase from about 2s to about 4s. Then, the observers' pupil responses were separated according to displayer's real and fake smile stimuli, from about 4s after the stimulus onset. These differences were observable clearly for male observers and averagely for female observers till the end of the viewing stimuli.

The outcomes of this experiment show that we can access the observer's different recognition and responses to real smile compared to fake smile stimuli via their pupil diameter. This is because the size of pupil is controlled by their autonomic nervous system which is known to respond to a person's emotional state [19]. According to evolutionary hypotheses of sex differences in emotional processing, the anatomy as well as function of the brain of the male is different (statistically) from the female [20]. Our results support the notion that pupil responses varied according to different emotional stimuli and have some similarity with recent findings that females respond more strongly to positive sounds and males respond more strongly to negative ones [21]. This is also similar to the average pupil size measured in [19] to discriminate negative from positive emotion. Furthermore, they found significant differences between emotions which are consistent with our outcomes considering fake versus real smile stimuli. We can also construct socio-biological explanations why males would respond more strongly to acted smiles while females would respond more strongly to real smiles.

The measurement of pupil size has important advantages compared to considering other physiological signal measurements. No sensors need be attached to the user, and pupil size may not be as sensitive to artefacts caused by various bodily movements. An important benefit of pupil size measurement is that it is not easy to control pupil size variation voluntarily and, thus, provides spontaneous and non-conscious outcomes. On the other hand, visually observable changes in facial expression can be masked, inhibited, exaggerated, and faked [2]. In this respect, pupil size measurement avoids these problems that are inherent in monitoring with video cameras. In this paper, we wanted to keep the video stimuli in a luminance range to make these as natural as possible. But careful attention is required to design a system in human-computer interaction for affective computing, because pupil size can vary according to light reflex, different stimulus parameters (e.g. visual and chemical), information-processing load and so on [12]. This control environment will be considered in our future research to compare with the present study and natural environments.

In conclusion, we analysed average pupil diameter timelines over 19 stimuli with 10 young adults from Asian backgrounds. These provide us an indication that the observer's pupil size varied differently for our fake smile stimuli as compared to real smile stimuli. Our outcome also provides evidence for gender differences. These findings suggest that by extracting and analysing an observer's pupillary responses, we can disambiguate real from fake smiles, based on the observer's innate and non-conscious recognition of the stimuli as representing fake or real smiles.

References

1. Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, **58**, 342–353.
2. Ekman, P., Friesen, W.V. (1982). Felt, false, and miserable smiles'. *Journal of Nonverbal Behavior*, **6**, 238–252.
3. Hoque, M., Morency, L. P., Picard, R. W., (2011). Are You Friendly or Just Polite? – Analysis of Smiles in Spontaneous Face-to-Face Interactions, Lecture Notes in Computer Science, 6974, 135-144.
4. Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, **115**, 268–287.
5. Valstar, M. F., Gunes H., Pantic, M. (2007). How to Distinguish Posed from Spontaneous Smiles using Geometric Features, Proceedings *of ACM International Conference on Multimodal Interfaces* (Nagoya, November 2007), 38-45.
6. Ochs, M., Niewiadmoski, R., Pelachaud, C. (2010). How a virtual agent should smile? Morphological and Dynamic Characteristics of virtual agent's smiles, *Proceedings of 10th International Conference on Intelligent Virtual Agents,* (Philadelphia, 20-22 September 2010), 427-440.
7. Ambadar, Z., Cohn, J. F., Reed, L. I. (2009). All Smiles are Not Created Equal: Morphology and Timing of Smiles Perceived as Amused, Polite, and Embarrassed/Nervous, *J Nonverbal Behav*. 33(1), 17–34.
8. Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, **91(2)**, 276-292.
9. Chapman, C. R., Oka, S., Bradshaw, D. H., Jacobson, R. C., Donaldson, G. W. (1999). Phasic pupil dilation response to noxious stimulation in normal volunteers: Relationship to brain evoked potentials and pain report. *Psychophysiology*, **36(1),** 44-52.
10. Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., et al. (2009). The Wick in the Candle of Learning: Epistemic Curiosity Activates Reward Circuitry and Enhances Memory, *Psychological Science*, **20**, 963-973.
11. Wang, J. T.-y., Spezio, M., Camerer, C. F. (2009). Pinocchio's Pupil: Using Eye tracking and Pupil Dilation To Understand Truth-telling and Deception in Sender-Receiver Games, *American Economic Review*, **100(3),** 984-1007.
12. Hess, E.H., Petrovich, S.B., (1987). Pupillary behaviour in communication, In: *Siegman*, A.W., Feldstein, S. (Eds.), Nonverbal Behaviour and Communication. Erlbaum, Hillsdale, NJ, 327–348.

13. Dhall, A., Goecke, R., Joshi, J., Sikka K., Gedeon, T. (2014). The Second Emotion Recognition In The Wild Challenge, *16th ACM Int'l Conf. on Multimodal Interaction*, (Istanbul, September 2014).
14. Pantic, M., Valstar, M., Rademaker, R., Maat, L. (2005). Web-Based Database for Facial Expression Analysis, *Proc. IEEE Int'l Conf. Multimedia and Expo*, (Netherlands, 6-8 July 2005), 5.
15. Soleymani, M., Lichtenauer, J., Pun, T., Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging, *IEEE Transactions on Affective Computing*, **3(1),** 42 – 55.
16. Petridisa, S., Martineza, B., Pantica, M. (2013). The MAHNOB Laughter database, *Image and Vision Computing*, **31(2),** 186–202.
17. Picard, R. W. (2000). Affective Computing, the MIT Press, Cambridge, Massachusetts, London, England, 175
18. Willenbockel, V., Sadr, J., Fiset, D., Horne, GO., Gosselin, F., Tanaka, JW. (2010). controlling low-level image properties: the SHINE toolbox, *Behav Res Methods*, **42(3),** 671-684.
19. Aracena, C., Basterrech, S., Snášel V., Velásquez, J. (2015). Neural Networks for emotion recognition based on eye tracking data, Proceedings of IEEE SMC Conference, (Hong Kong, 9-12 October 2015).
20. Cahill, L. (2005). His Brain, her brain, *Scientific American*, **292**, 40-47.
21. Partala T., Surakka, V. (2003). Pupil size variation as an indication of affective processing, *International Journal of Human-Computer Studies*, **59(1-2),** 185-198.
22. Blair, R. C., Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials, *Psychophysiology*, 30, 518-524.

Verification of Autonomous Vehicle Over-reliance

Toshiya Arakawa[1] and Kunihiko Oi[2]

[1]Department of Mechanical Systems Engineering, Aichi University of Technology, Aichi, Japan.
arakawa-toshiya@aut.ac.jp
[2]Toyohashi Sekkei Inc., Aichi, Japan.

1. Introduction

Recently, there is a growing interest in developing autonomous vehicles by automakers and others. In 1939, at the New York World's fair General Motors (GM) hosted an exhibit called "Futurama." GM envisioned a future in which many autonomous vehicles would drive in a city with a lot of green areas. Hence, the idea of an autonomous vehicle is not recent. Several attempts were made to realize the idea of an autonomous vehicle. These included the development of autonomous vehicles such as Stanford Cart, Boss by Carnegie Mellon University, Shelly by Audi TTS, and Toyota Prius by Google. Additionally, an increasing number of automakers and others are also unveiling their plans. In December 2013, Ford unveiled a self-driving research car called a modified Fusion Hybrid. In 2017 in Gothenburg, Sweden, Volvo announced the introduction of 100 autonomous vehicles. Google has a fleet of fully autonomous Priuses and will release the corresponding technology in 2018. Nissan plans to announce the release of the first "commercially viable" self-driving system by 2020. GM, Audi, BMW, Tesla and others also have autonomous vehicles under development. The 2014 Mercedes-Benz S-Class was the first car in the market that was fully capable of driving itself. With these developments, the idea of a fully autonomous vehicle and not just a partial autonomous vehicle like a vehicle equipped with ACC (adaptive cruise control) and/or LKAS (lane keeping assist system), may come true. An HIS automotive study predicted that there will be nearly 54 million self-driving cars in the world by 2035. The study also predicted that after 2050, nearly all vehicles in use are likely to be self-driving cars or self-driving commercial vehicles. As an increasing number of autonomous vehicles are developed, the relation between the vehicle and driver will become more complex. The autonomous vehicle technology could be viewed by using a five-part continuum as suggested by the National Highway Traffic Safety Administration (NHTSA), with different benefits realized at different levels of automation [1][2](Table 1).

Table 1. Automation levels by NHTSA.

Level 0	The human driver is in complete control of all functions of the car
Level 1	One function is automated
Level 2	More than one function is automated at the same time, but the driver remains attentive.
Level 3	Driving functions are sufficiently automated–the driver can safely engage in other activities
Level 4	The car is self-driving–no human driver required

As the autonomous vehicle is a very feasible technology, people expect a mass production of autonomous vehicles. However, several problems need to be solved in order for autonomous vehicles to be produced. These problems include law revision, infrastructure development of P2P (Person to Person), V2P (Vehicle to Person), V2I (Vehicle to Infrastructure), and V2V (Vehicle to Vehicle) communication. Additionally, as the current focus of developers includes realizing level 3 of autonomous vehicles, human factors should be adequately considered with respect to driving autonomous vehicles. A study indicated that drivers were incentivized and their mental workload (such as the effort in extending an arm) was reduced when ACC (Adaptive Cruise Control) was used.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

It also suggested that the drivers tend to rely heavily on the system [3]. However, the driver could also rely on system excessively [3]. Thus, it is very important to verify whether or not a driver relies excessively on the autonomous vehicle. The manner in which the driver handles a normal situation or relies on the autonomous vehicle excessively is also important. If the results reveal a tendency to excessively rely on autonomous vehicles, then this could be recognized as a negative effect of autonomous vehicles, and mechanisms to deal with the same can be investigated.

This study analyzes whether or not drivers rely on autonomous vehicle excessively from the human factor viewpoint. This paper is divided as follows. The experimental situation and methods are introduced in Section 2. Section 3 discusses the results of the experiment. In section 4, the analyzed results are demonstrated. Section 5 consists of the conclusions.

Ethical statement: It should be noted that in this study all procedures were conducted in accordance with the research ethics guidelines at the Aichi University of Technology.

2. Methods/Experiments

Three subjects (aged 20–21 years) were involved in the experiment. A driving simulator (D3sim, by Mitsubishi precision Co., Ltd.) was used to measure the effect of human factors on autonomous driving. Drivers were instructed to drive on the urban-area course for approximately fifteen minutes by controlling the handle, accelerator pedal and braking pedal on their own. This was the manual-driving scenario. After a break of approximately 15 min, the drivers were instructed to drive on the urban-area course for another period of approximately 15 min. However, in this session, the handle, accelerator pedal and braking pedal were automatically controlled. Thus, the drivers did not have to control the vehicle themselves. This was the autonomous driving scenario. Drivers were also asked to press the brake pedal when a presence of the collision risk was determined, and/or the system fault occurred suddenly, and/or the drivers could not control the driving simulator. It should be noted that in reality, no system fault occurred. However, in the autonomous driving scenario, an event where a man suddenly rushed out on the course after approximately 6 min of driving was included.

In both the manual and autonomous driving scenarios, the driver's eye movement, facial expression, braking behavior and experimental situation were recorded and measured. The driver's eye movements were recorded and measured by an eye-mark recorder (T.K.K.2950 TalkEye Lite, by Takei Scientific Instruments Co., Ltd.). The driver's facial expression and braking behavior were recorded as motion pictures by a CCD camera and a small infrared camera, respectively. The experimental situation was recorded by a video camera (Everio, by JVC). These scenes were combined into a single screen by a split screen unit (Figure 1).

The timing of attaching a foot on the braking pedal or hitting the brakes was judged by the experimenter from the motion picture of driver's facial expression. With respect to the time-series data on the sleep level of the driver, the sleep level of the driver was based on the NEDO index. The NEDO index is the method of determining sleep levels in which individuals observe a motion picture of the driver, and the sleep level of the driver is judged by a majority vote. The sleep level based on the NEDO index was categorized as follows [4]:

(1) Sleep level 1: not at all sleepy (the subject's eye-movements are fast and frequent).
(2) Sleep level 2: slightly sleepy (the subject's eye-movements are slow and the mouth is open).
(3) Sleep level 3: sleepy (the subject's eye blinks are slow and frequent, the mouth moves, and the subject corrects his/her posture).
(4) Sleep level 4: pretty sleepy (the subject displays a conscious eye blink, the subject shakes his/her head, the subject yawns frequently).
(5) Sleep level 5: very sleepy (the subject closes his/her eye, the subject leans his/her head back and forth).

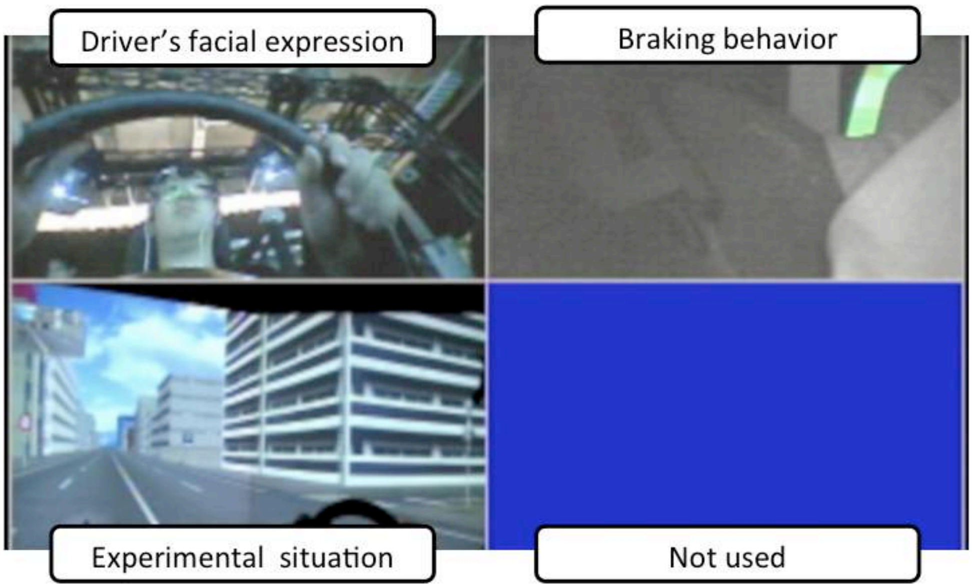


Figure 1. Driver's facial expression (upper left), braking behavior (upper right) and experimental situation (lower left).

3. Results

The heat map (that is, time-series data and probability of eye-movement that indicates where a driver tends to gaze), the time-series graph of autonomous braking, the driver's brake work, and time-series data of the driver's sleep level were illustrated in Fig. 2, Fig.3 and Fig. 4 for participants A, B, and C, respectively. The heat map shows where and how long the driver gazed in the driving direction. The horizontal direction was the angle of the gazing direction, wherein the center was 0°, the left end of the field of vision was − 90°, and the right end of the field of vision was 90°. The vertical direction indicated the driving time from the start of the vehicle. The heat map illustrated the probability density of the gazing point in a time series. If an area in heat map was red colored at a point, it implied that the probability of the driver's gaze in the area was relatively high. Conversely, if an area in the heat map was blue colored at a point, this indicated that the probability of the driver's gaze in the area was relatively low. In the time-series graph of autonomous braking and the driver's brake work, the blue line depicted the timing of autonomous braking, and the red line represented the timing of attaching a foot on the braking pedal or hitting the brakes judging from NEDO index.

4. Discussion

Figure 2(A) and figure 3(A) indicated that participants A and B mainly looked in the driving direction, and they appeared to concentrate on the safety of driving in the case of the manual driving scenario. Figures 2(B) and 3(B) indicated that drivers tended to look in the driving direction, but they also looked at the areas that were unrelated to driving. For example, in figure 2(B), the participant A tended to look in the area between − 80° and − 60° with time. From figure 3(B), participant B tended to look in the right from the center, and the frequency of looking at the 60° area increased. This suggested that participants A and B were slightly nervous about driving in the manual-driving scenario, but their mental workloads were decreased. Thus, they could look not only in the driving direction but also in the areas that were unrelated to driving. This tendency was caused by a reliance on the autonomous system. If the drivers did not over rely on the autonomous system, they looked only in the driving direction. They looked at the areas that were unrelated to driving only after spending a long time driving. However, in this study, the subjects looked at the areas that were unrelated to driving as soon as they started driving. Hence, it could be concluded that the subjects relied heavily on autonomous driving.

From figure 2(C) and figure 3(C), drivers tended to not attach a foot on the braking pedal with time. If drivers did not rely on the autonomous vehicle, they attached a foot on the braking pedal constantly, and hit the brake every time that the autonomous braking was in progress. The subjects in the study rarely placed a foot on the braking pedal. This suggested that they seemed to rely on the autonomous vehicle. Participant A and B did not attach a foot to the braking pedal, and were therefore not prepared for the scenario when a man rushed out. The participants assumed that the autonomous system would avoid the obstacles, and would always stop the vehicle before it collided with the obstacles. Thus, the participants relied on the autonomous driving to the extent that they were not prepared by attaching a foot on the braking pedal, in spite of being warned about the possibility of system failure.

Figures 2(D) and 3(D) indicated that the sleep levels of all participants increased gradually. For instance sleep level at the start of the driving scenarios was one, while in contrast, the sleep level was two or three at the end of driving. Thus, the mental workload of the subject was very low in the autonomous driving scenario, and the subject did not feel the tension at all. It was observed from the heat map, the braking behavior, and the change of sleep levels that participant A relied too heavily on the autonomous vehicle. Conversely, from figure 4(A) it was revealed that participant C looked in the driving direction in both the manual driving and the autonomous driving scenarios, and s/he did not display a tendency to look at the areas that were unrelated to driving. Although participant C had a license, the participant rarely drove. This limited his/her driving experience. Additionally, the subject was afraid that the system fault might occur suddenly. Thus, the subject concentrated on driving excessively. Hence, taking these factors into account, the subject's mental workload was very high and s/he did not look at the areas that were unrelated to driving. In fact, after the subject completed the driving task, s/he revealed that s/he concentrated on driving because of the fear that the system could fail. This was supported by figure 4(B). In figure 4(B), almost all the blue lines and the red lines overlapped, suggesting that the subject did not rely on autonomous driving, and that s/he had a foot attached on a braking pedal throughout the action. However, from figure 4(C), it was suggested that participant C also became sleepy, as his/her sleep level was three at the end of the driving task. Thus, participant C concentrated too much on driving but s/he could not control vehicle. That is, the participant concentrated excessively and therefore became bored and sleepy.

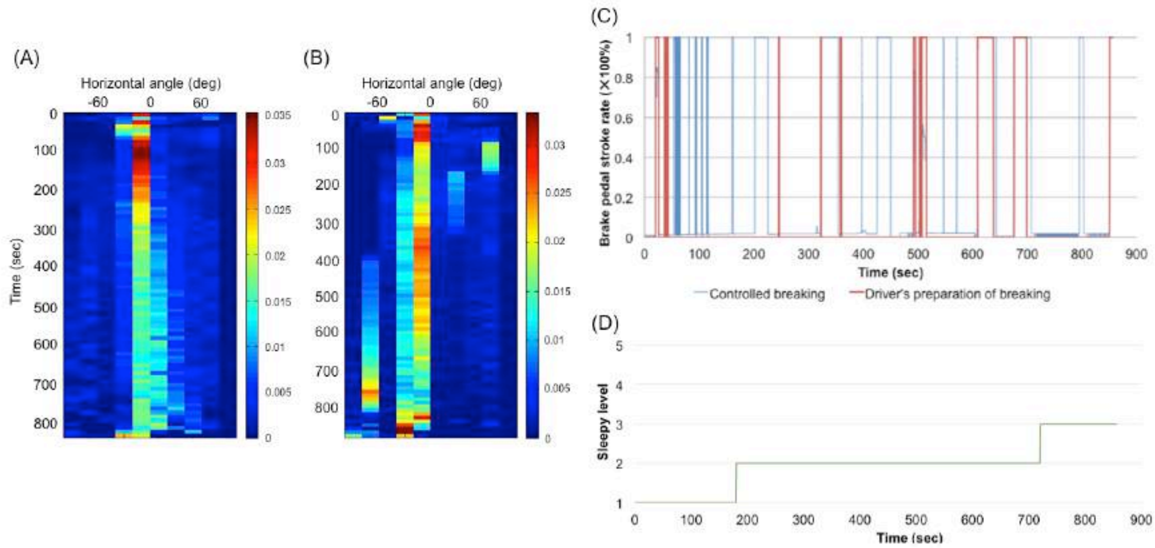


Figure 2. Heat map of manual driving(A), heat map of autonomous driving(B), the time-series graph of autonomous braking and driver's brake work(C), sleep level (D) of participant A.

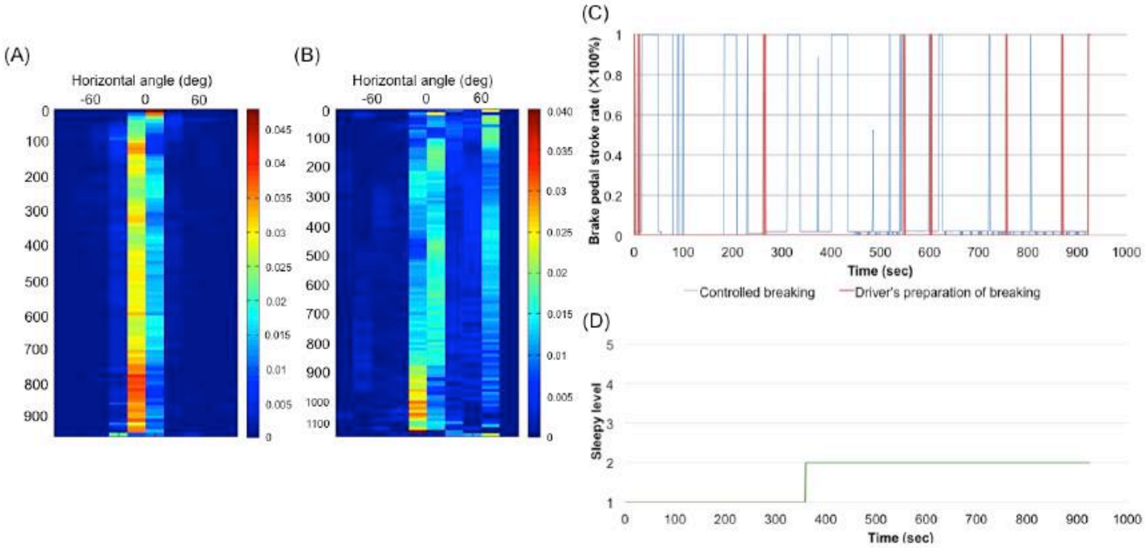


Figure 3. Heat map of manual driving (A), heat map of autonomous driving (B), the time-series graph of autonomous braking and driver's brake work (C) and sleep level (D) of participant B.

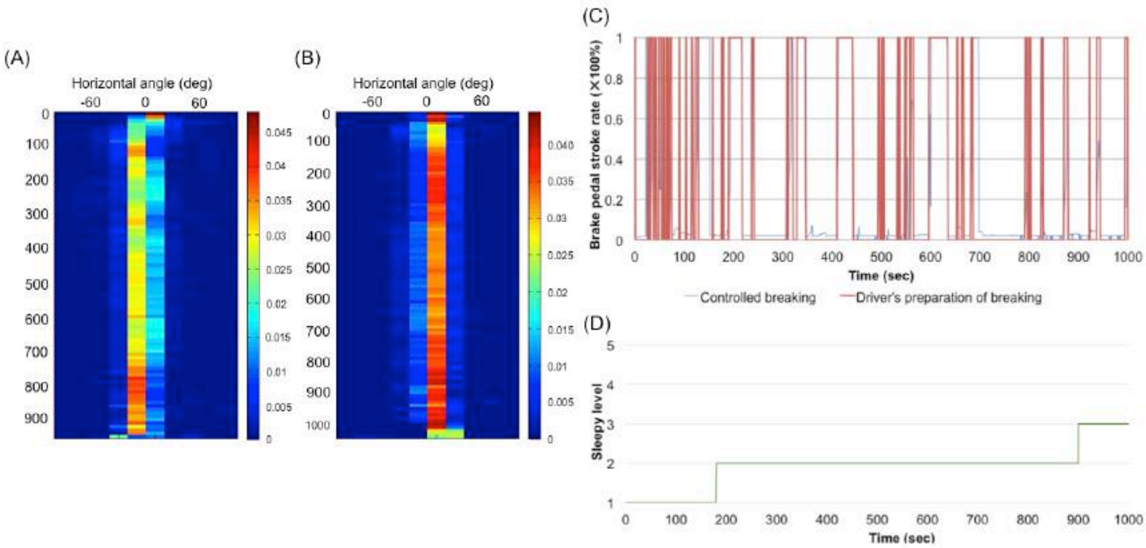


Figure 4. Heat map of manual driving (A), heat map of autonomous driving (B), the time-series graph of autonomous braking and driver's brake work(C) and sleep level (D) of participant C.

5. Conclusion

This paper verified the excessive reliance of drivers on an autonomous system or autonomous vehicle. From the eye-movement viewpoint, drivers gazed at objects unrelated to driving, such as side strips and roadside trees. Additionally, drivers tended not to attach a foot on the braking pedal as they relied on the autonomous system, and believed that the autonomous vehicle would stop and not hit the obstacle. Thus, the mental workload of driver in the autonomous system driving scenario was less than the mental workload of the driver in the manual driving scenario. Finally, the sleep level of the driver based on the NEDO index verified the driver's reliance on the autonomous system. It was found that drivers became sleepy while driving autonomous vehicles.

This paper does not suggest that autonomous vehicles should not be used given the tendency of drivers to rely on autonomous vehicles. It is true that autonomous vehicles help in improving the driver's quality of life (QOL). Hence, we stress on the importance of supplementing the tendency to rely on the autonomous system or vehicle with additional manual and visual control, and alertness.

Acknowledgements

This research was supported by KAKENHI (Grant-in-Aid for Scientific Research) 15K21486.

References

1. Frank, D. (2014). Autonomous Vehicles: The Automotive Ecosystem. <https://frankdiana.wordpress.com/2014/03/03/autonomous-vehicles-the-automotive-ecosystem/>. Accessed 22 January 2015.
2. National Highway Traffic Safety Administration. (2013). Preliminary Statement of Policy Concerning Automated Vehicles. <http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf>. Accessed 22 January 2015.
3. Inagaki, T. (2010). Human's Overtrust in and Overreliance on Driver Assistance Systems. *ICICE Technical Report* SSI2010-6, 21-24.
4. Yasuhiko, N., Ayu, M., Satoshi, S. (2008). Detection of Driver Drowsiness Level. *Fujitsu* 59, 4, 416-420.

An Information-Theoretic Approach to Comparing Time Series of Driving Behaviour between Healthy and Glaucoma Drivers

Erwin R. Boer, Alberto Diniz-Filho and Felipe A. Medeiros

Laboratory of Performance and Visual Function, Department of Ophthalmology, University of California San Diego, La Jolla, California. eBoer@ucsd.edu, aDinizFilho@ucsd.edu, fMedeiros@ucsd.edu

Introduction

Glaucoma is an eye disease that degrades a person's low level visual processing ability; it increases the time it takes to process visual information, it degrades a person's ability to detect objects under low luminance/contrast conditions, and it may create blind spots (scotomas) across the peripheral visual field [10]. Early diagnosis aids in slowing disease progression while functional assessment aids in counselling patients about the increased risks they face in daily life; driving being an important one. In this study we focus on the development and evaluation of a sensitive metric that can quantify whether glaucoma patients are exposed to greater driving risk than their healthy counterparts. The test conditions are simulated low contrast and limited visibility driving down a winding country road in our full scale motion-enabled driving simulator.

Comparison of time series that capture a human operator's behaviour are generally based on aggregate performance metrics obtained over some portion of the time series or at specific points within the time series. These time series can be human operator control actions, controlled vehicle states, eye movements, voice intonations during the reading of a passage, etc. The example, addressed in this paper, is a comparison of the risk experienced by healthy and glaucoma drivers while driving a winding road. Typical comparisons are based on the following metrics: i) standard deviation of lateral position within the lane [7], ii) number of steering reversals [8] or iii) minimum time to line crossing [4,5,6]. Each of these metrics has its own merits but also shortcomings.

On a straight road, the standard deviation of lateral position (SDLP) is a reasonable *performance* metric except that drivers are satisficing controllers [3] and therefore will accept a range of lateral positions especially when heading parallel to the lane boundaries. This would place a threshold on SDLP that upper bounds all equally acceptable lateral positions thus reducing the metric's sensitivity. On a winding road, natural curve cutting behaviour would inflate SDLP. It can be argued that more curve cutting equates better performance even though it would increase SDLP. Furthermore, SDLP does not distinguish between taking a curve wide or even over shooting it and cutting it; clearly two situations with very different risks. Thus, individual differences in the acceptable range of lateral positions and differences in curve-cutting strategies would increase the variability in SDLP across subjects, reducing its sensitivity to quantify behavioural differences between conditions of interest.

Because drivers try to protect self-imposed or experimentally-imposed safety margins, time to line crossing (TLC) has been used as a meaningful metric to quantify the *risk* drivers are exposed to [4,5,6]. TLC is the time it would take the vehicle to leave the lane if the current heading were maintained (straight TLC) or if the current yaw rate or equivalently steering angle were maintained (curved TLC). By taking the minimum value, a single risk point is obtained that maybe in a specific tight curve and thus similar for all drivers or simply a single attention lapse across the drive. Minimum TLC therefore either ignores a large amount of valuable data or does not capture the difference between drivers who have many short TLCs and those that only have a few. It is also too sensitive to random fluctuations in attention. Even with these limitations, researchers often do take the minimum TLC instead of other statistics such as mean or standard deviation for several reasons. On a straight road, TLC can be very large or infinite (numerical singularity) when the vehicle is heading parallel to the lane boundaries thus making the mean TLC a value that is only minimally affected by the few short TLCs that drivers exhibit. Relatedly, the distribution of TLC values along a curved road is also highly skewed thus making it necessary to either take the minimum TLC or some low percentile TLC value. People have used inverse TLC

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

(InvTLC) to avoid the numerical singularity and producing a less skewed distribution but the lack of sensitivity to one or more critically small TLCs remains limited.

Number of steering reversals or steering reversal rate (SRR) is a different type of metric that captures the amount of *effort* that drivers expend to maintain their vehicle within acceptable safety margins [8]. At slow speeds on a winding road drivers will show steering reversals that are related mostly to maintain the vehicle within an acceptable range of lateral positions because the experienced risk is generally continually within their acceptable range (i.e. TLC is long enough). At higher speeds drivers begin to cut curves and try to adopt the optimal race-line to maximize their TLC which reduces the number of steering reversals in the limit down to the number of curves and thus the curvature profile begins to dominate the metric. Thus steering reversals may not be a good metric because it depends too much on speed and the adopted control strategy.

As is the case with most metrics, when the effect of the studied conditions is large enough, the metrics will become significantly different. While this may be acceptable for many studies, in the medical field the sensitivity of a diagnostic test has a strong impact on the adopted treatment and generally early onset of treatment leads to better outcome in terms of limiting the impact that a disease has on a person's life long term. The goal in this study was to explore whether a more sophisticated approach that is grounded in information theory can yield a substantial improvement in diagnostic strength. The new behavioural entropy metric detailed below will be compared against the SDLP and mean InvTLC discussed above to assess whether it is capable of more sensitively capturing differences in behaviour between healthy subjects and glaucoma patients.

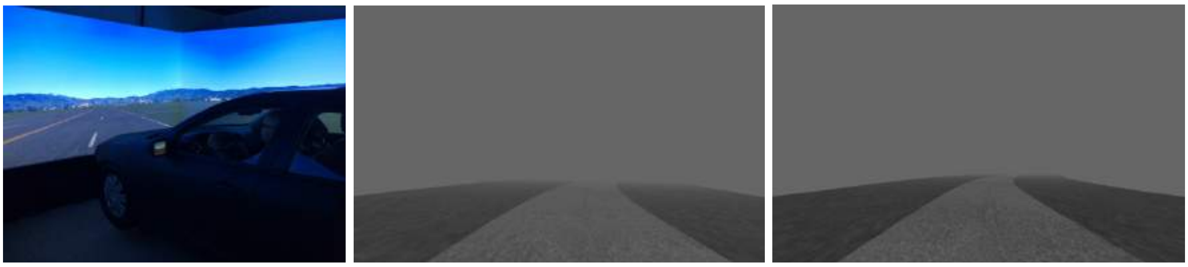


Figure 1. Left panel shows a picture of the driving simulator at the Hamilton Glaucoma Institute at UCSD. It provides a realistic driving experience with simulated vehicle motion and vibration, natural steering and pedal feel, and a 180 degree horizontal field of view plus rear and side mirror views. The provider is Realtime Technologies, Inc. (http://www.simcreator.com/). The middle panel shows the experimentally used front scene image at 30m visibility and the right panel the 60m visibility. The contrast between the light lane and the darker ground surface was 25%.

Test Description

Healthy subjects and glaucoma patients each drove a 2.82km long winding road in a driving simulator under a short (30m) and long (60m) visibility conditions (Figure 1) while their speed was controlled automatically from slow (15mps) to medium (20mps) to fast (25mps) as shown in Table 1. The curvature fluctuated according to a sum of sinusoids such that the minimum radius was about 150m and therefore the lateral accelerations drivers experienced in each of the three were approximately 1.5g, 2.7g and 4.2g respectively. The study included 41 glaucoma patients and 25 healthy control subjects whose demographics are shown in Table 2.

The test instruction was to stay within the lane. Participants from this study were included in a prospective longitudinal study designed to evaluate functional impairment in glaucoma conducted at the Laboratory of Performance and Visual Function of the University of California San Diego. The institutional review board at the University of California San Diego approved the methods, and written informed consent was obtained from all participants. The study adhered to the laws of the Health Insurance Portability and Accountability Act, and all study methods complied with the Declaration of Helsinki guidelines for human subject research.

Entropy Calculation

The goal is to develop a metric that sensitively captures difference in driving behaviour between glaucoma patients and healthy subjects. Because we are concerned about our patients' driving safety, we use InvTLC rather than lateral position as the instantaneous risk experienced at every time step. Thus for each driver we have per speed segment a time series of straight InvTLC values. An arguably better measure would be the curved InvTLC but since the focus of this paper is to introduce and explore the utility of a behavioural-entropy metric we selected the measure that is most easily calculated from the raw driving simulator data. Figure 2 shows an example of straight TLC when a vehicle slightly overshoots a sharp curve.

Table 1. Experimental visibility conditions for three speed segments along the 2820m winding road.

Experimental Condition	Visibility		Speed	
		Slow 15mps Start/End Segment [m]	Medium 20mps Start/End Segment [m]	Fast 25mps Start/End Segment [m]
V30	30m	0,920	920-1920	1920-2820
V60	60m	0,920	920-1920	1920-2820

Table 2. Demographic and clinical characteristics of subjects included in the study

	Glaucoma (n = 41)	Control (n = 25)	P-value
Age, years	69.0 ± 10.9	65.7 ± 9.9	0.220[a]
Gender, n (%) female	9 (22.0)	12 (48.0)	0.033[b]
Race, n (%)			
White	20 (48.8)	15 (60.0)	0.163[b]
African-American	9 (21.9)	9 (36.0)	
Asian	8 (19.5)	1 (4.0)	
Other	4 (9.8)	0 (0.0)	
Binocular MS SAP 24-2, dB	24.2 ± 7.4	31.2 ± 1.7	<0.001[c]
MD SAP 24-2 (worse eye), dB	-12.8 ± 10.3	-0.7 ± 1.5	<0.001[c]
MD SAP 24-2 (better eye), dB	-6.5 ± 7.2	0.2 ± 1.2	<0.001[c]
Visual acuity (worse eye), logMAR	0.11 ± 0.23	-0.01 ± 0.14	0.004[c]
Visual acuity (better eye), logMAR	0.02 ± 0.12	-0.07 ± 0.12	0.004[a]
Average mileage driven, miles/week	116.0 ± 107.5	117.6 ± 116.5	0.942[c]
Montreal Cognitive Assessment score	28.0 ± 2.2	28.3 ± 2.5	0.415[c]

MD = mean deviation; SAP = standard automated perimetry; dB = decibels; MS = mean sensitivity; logMAR = logarithm of the minimum angle of resolution. [a]Student t test. [b]Fisher's exact test. [c]Wilcoxon rank-sum test.

Some of the behavioural entropy metrics developed by the lead author, most notably steering entropy, have been recognized as one of the more sensitive ways to quantify differences in performance and workload [1,2,7]. These entropy metrics were based on distributions of steering angle prediction errors accumulated across the entire drive. In other words, the baseline distributions were established based on whole drive data. Here we establish a metrics where a baseline distribution is established at every location down the road.

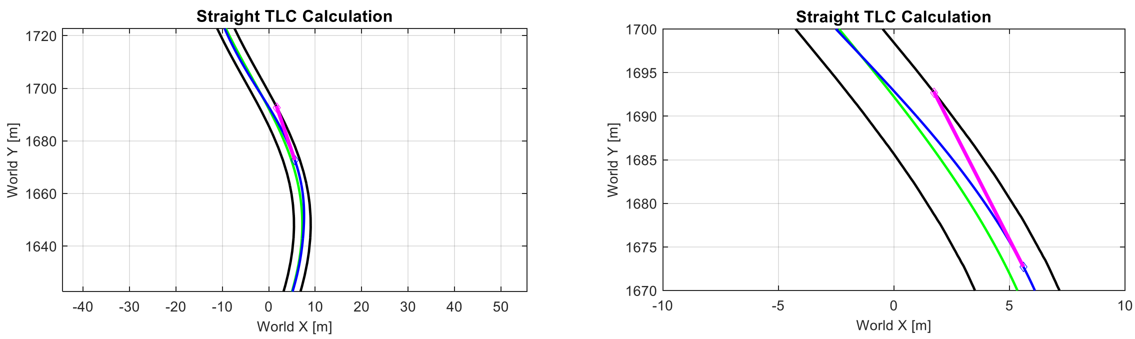


Figure 2. Straight TLC as experienced by one of the participants overshooting a left curve. The green line indicates the centre of the lane demarked by black lane boundary lines. The blue line is the vehicle trajectory and the magenta line show the straight heading vector intersecting with the right lane edge where the straight TLC point is defined.

It is expected that sensitivity and utility of a metric are increased when it satisfies the following criteria applied to the case at hand: i) the metric is independent of the adopted control strategy by healthy subjects but sensitive to changes in control strategy as a result of the disease, ii) the data should be normalized to that of healthy subjects not just at the aggregate level but at the level of each sample (i.e. location down the road), iii) the comparison between glaucoma patients and healthy subjects should be independent of the shape of the raw data distribution (i.e. direct probability calculation), iv) the metric should not be plagued by tuneable parameters such as percentiles or thresholds (i.e. parameter free metric), and v) the metric should be generalizable to other continuous driving tasks such as car following where inverse time to collision (InvTTC) and be used and lateral disturbance fighting on a straight road where InvTLC can also be used (i.e. generalizable). The adopted metric is called Behavioural Entropy and is detailed below in two subsections: data preparation and entropy calculation.

The following steps describe the ***data preparation***. This preparation is done for each participant's drive in both visibility conditions. The algorithm is detailed around the case at hand to hopefully make it easier to understand than a purely abstract exposition. Each driver drove the same winding road twice: once with low 30m visibility and once with higher 60m visibility as depicted in Figure 1.

1. The time series data is limited to within the valid x-range (i.e. from start of slow speed range to end of fast speed range as detailed in Table 1).

2. The straight InvTLC τ^{-1} is computed for every time sample by simply finding the intersection point between the vehicle heading vector and a lane boundary. Outlier cases are discussed in steps 8-12.

3. The InvTLC time series is discretized into reference road cells that are as long as the vehicle travels within one time step (16.67ms). Because the speed increases twice down the road, the road cells in the slow segment are shorter than those in the fast section. These reference road cells are the same for all drivers so that data can be accumulated and compared per road point. The calculated InvTLC values are put in the nearest road cell. In all but the most extreme cases such as when a driver substantially leaves the lane, each reference road cell n receives exactly one InvTLC value for each subject k. See Figure 3 for details on the mapping from consecutive vehicle trajectory samples to reference road points.

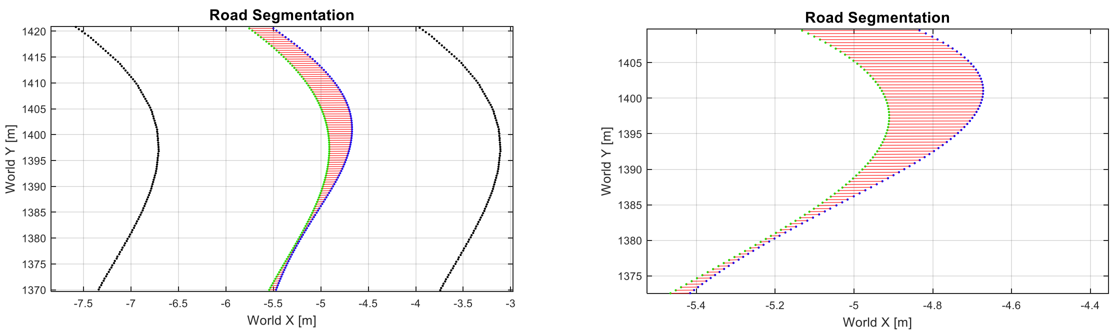


Figure 3. The road is segmented with a spacing that equals the average distance travelled within 16.67ms (60Hz update rate in the driving simulator). This means that consecutive samples in the vehicle trajectory (blue dots) will project to consecutive samples in the segmented road profile (green dots along centre line). The two black dotted lines denote the lane boundaries.

The following step is only applied to all of the healthy subject's data because it produces the baseline against which the glaucoma patients' data is compared.

4. Combine all InvTLC road cell data from all healthy subjects K_H across both conditions (*V30+V60*). This leads to a data set that has twice as many InvTLC values as there are healthy subjects in each road cell. These data form the baseline InvTLC distributions at every road point against which the individual road point InvTLCs from glaucoma patients are compared in an information theoretic fashion as detailed below. The data from individual healthy subjects is also compared against this baseline to provide an indication which healthy subjects are outliers within the healthy subject group.

The set of InvTLC data τ^{-1} per road point $n_{\{s,m,f\}}$ from all healthy subjects' $\{K_H\}$ across both visibility conditions *V30* and *V60* is referred to as the baseline cumulative distribution function (CDF) for a sample in the slow n_s, medium n_m or fast n_f section of the road. Mathematically this baseline CDF is referenced as:

$$ {}^{V30+60}_{\{K_H\}}CDF_{n_{\{s,m,f\}}}\left(\tau^{-1}_{n_{\{s,m,f\}}}\right)=\left\{ {}^{Vm}_{k_H}\tau^{-1}_{n_{\{s,m,f\}}} \middle| k_H\in[1,K_H],m\in\{30,60\}\right\}. $$

For clarity, the sub and super-scripts around different variables all adhere to the following nomenclature: ${}^{\text{Condition(s)}}_{\text{Participant(s)}}X_{\text{RoadPoint(s)}}$.

The observed InvTLC time series mapped to reference road points for all healthy subjects and all glaucoma patients for the two visibility conditions are shown in Figure 4. It shows clearly that the spread of InvTLC values at corresponding road points is generally greater for glaucoma patients than for healthy subjects. The valleys in the InvTLC profiles are at points where the road transitions from a left to a right curve or vice versa and thus the TLC is temporarily longer (InvTLC shorter). The peaks in the InvTLC profile are around the apex of the curves. The InvTLC values are capped at $10s^{-1}$ (TLC = 0.1s) for reasons detailed in steps 8 through 12.

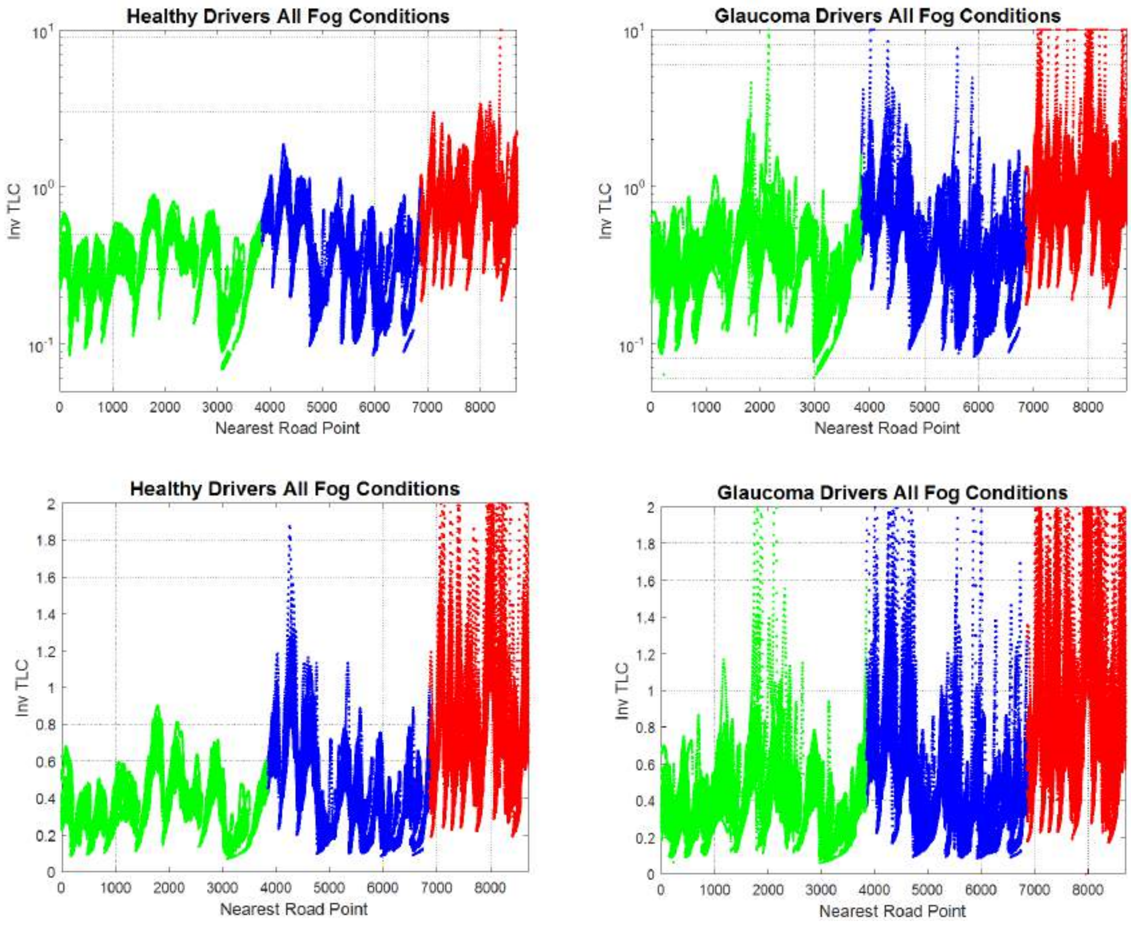


Figure 4. Left panels show InvTLC for all healthy subjects in both visibility conditions *V30* and *V60*. The green, blue and red colours refer to the slow, medium and fast speed segments respectively. The right panels show the same for glaucoma patients. The top row shows InvTLC values along a log scale while the bottom road shows the same data in a linear scale.

Now we have a distribution of InvTLC values assembled from all healthy subjects' data at every road point. This enables us to take an InvTLC value from a glaucoma patient at any road point and compute how likely it is that such an InvTLC would have been produced by a healthy subject at that location down the road. If it is unlikely then we assign a high value to that sample, and conversely if it is likely then we assign a small value. This likelihood based assignment is based on information theoretical principles and is detailed next.

The following steps detail the ***entropy calculation***. For each observed InvTLC at some road point from either a healthy subject or a glaucoma patient at some speed in some visibility condition the following set of steps are used to compute the individual sample bit value as well as the entropy of the data sub-set of interest (e.g. glaucoma at slow speed in 30m visibility).

5. The primed CDF probability of the InvTLC samples from either condition or group is computed using the baseline CDF at the samples road point. For glaucoma patient k_G in the *V30* condition at road point n_f we obtain

$$ {}_{k_G}^{V30}p'_{n_f} = 1 - 2\left| {}_{\{K_H\}}^{V30+60}CDF_{n_f}\left({}_{k_G}^{V30}\tau_{n_f}^{-1}\right) - 1/2 \right| $$

   where the prime is used because it is not the traditional CDF probability p_{CDF} but is computed as follows $p' = 1 - 2\left|p_{CDF} - 1/2\right|$; it is roughly related to the likelihood of the sample given the CDF distribution. This assures that the medium value where $p_{CDF} = 1/2$ results in $p' = 1$. Of course many other mappings from p_{CDF} to p' are possible but this one is easily computed and does not require a large number of data points to make up the CDF nor require fitting a distribution to the CDF.

6. The primed CDF probability is converted to the optimal number of bits required to represent it by simply taking the log base 2 of the primed CDF probability: ${}_{k_G}^{V30}h_{n_f} = -\log_2\left({}_{k_G}^{V30}p'_{n_f}\right)$. We refer to ${}_{k_G}^{V30}h_{n_f}$ as the *bit value* of glaucoma patient k_G's experienced risk at road point n_f under condition *V30.* Note that this the information theoretic value defined by Shannon and Weaver in 1949 [12].

7. These bit values can be combined into an entropy-metric that is computed over a specific portion of the data such as a particular speed or visibility condition. The entropy of glaucoma patient $k_G \in [1, K_G]$ over the slow section $n_s \in \{N_{slow}\} = [1, N_{slow}]$ in the long visibility condition (*V60*) is:

$$ {}_{k_G}^{V60}H_s = \sum_{n \in \{N_{slow}\}} {}_{k_G}^{V60}h_n $$

   Of course it is also possible to compute the entropy for a participant across all three speed segments. For example, the entropy of healthy subject k_H for the short visibility condition *V30* is computed according to ${}_{k_H}^{V30}H_{\{s,m,f\}} = \sum_{n \in \{N_{slow}, N_{med}, N_{fast}\}} {}_{k_H}^{V30}h_n$.

As mentioned, the normalization is based on the combined set of InvTLCs observed across all healthy subjects across both visibility conditions (*V30* and *V60*). It is of course possible to establish the baseline CDFs based on the long or the short visibility condition. For the behavioural entropy metric to be maximally sensitive, the data set used as baseline (i.e. normalization) should be data based on normal behaviour so that outlier behaviour becomes unlikely and thus receives many high bit values that inflate the entropy. In other words, the goal is to assure that one or more data sets (group and/or condition) yields many observations that are unlikely within the baseline data set because that yields high entropy values; to achieve this, the baseline data set should be chosen to only have a limited number of outliers. In our case, the baseline dataset would be chosen from healthy subjects in the long or short plus long visibility conditions. Here we chose the combination of short and long (i.e. all healthy subject data combined but could also have chosen just the 30m visibility data) again because the goal is to demonstrate the sensitivity of the behavioural entropy metric and not to explore many possible sensitivity improvement opportunities; those will be detailed in a forthcoming journal paper.

Implementation of Entropy Calculation

Steps 1 through 7 detailed in the previous section establish how to compute the baseline CDFs as well as to compute the entropy for a set of data. While these steps are theoretically correct, when it comes to implementation, several aspects need to be taken into consideration:

8. The InvTLC is computed based on the centre of the vehicle. This means that the left wheels are allowed to leave the lane as long as the centre of gravity of the vehicle does not. Of course different versions of TLC can be adopted (e.g. based on front left and right corners of the vehicle), but since the focus here is the method, these other types of TLC were not be explored.

9. InvTLC is based on the straight heading TLC. As mentioned before, curved TLC can be used instead and is arguably a better choice because it captures the time until the driver needs to make a steering change. Again, because the focus is on methodology, these alternatives will not be explored.

10. The InvTLC values should be upper limited because in a few cases participants drive out of the lane and those events would result in an infinite InvTLC because TLC goes through zero when crossing a lane boundary. The InvTLC upper limit is set to $10s^{-1}$ (equivalently, the minimum TLC is set to $0.1s$).

11. The baseline CDF is generally narrower than the range of InvTLC values observed in the non-baseline data. This would mean that the primed CDF probability would be zero in many cases which would yield an undefined bit value and thus an undefined entropy value. To remedy this, each baseline CDF of InvTLCs is expanded with one zero-value as well as one max-value of 10.0. The raw and expanded baseline CDFs are shown in Figure 5.

12. The baseline CDF at each road point is represented by the set of individual InvTLC values observed within the baseline data set at that road point; not some distribution parameterization to assure that the computed probabilities are independent of the underlying distribution. This means that the primed CDF probability needs to be computed using interpolation or nearest neighbour. We chose interpolation to assure as less discrete set of probabilities for outlier InvTLC values at a specific road point that fall outside the observed range within the baseline data set at the same road point. In case the InvTLC is 10, a CDF probability (note, not the primed CDF probability) of $1-1/M$ where M is the number of points in the baseline distribution (e.g. $M=2K_H$ because data from V30 and V60 from all healthy subjects was combined to create the baseline).

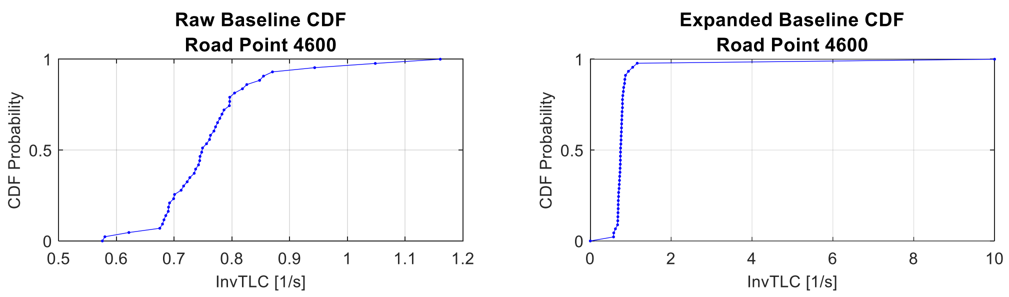


Figure 5. Raw (left panel) and expanded (right panel) baseline CDFs of InvTLCs (i.e. healthy subjects across both visibility conditions).

The behavioural entropy algorithm above is detailed for time series data where every subject performs the same task so that location specific data can be accumulated and location specific comparisons can be made. This also makes it possible to assess where along a road, for example, we observe the greatest effect which will aid in specific coaching of the patients. This location specific approach requires that a distribution (CDF) of some behaviour or measure at a particular location can be established. Note that exploration of a visual scene produces location (region) specific distributions of eye movement parameters such as glace duration, revisit time intervals, location of previous or subsequent fixation location. The behavioural entropy approach naturally applies to this type of eye movement behaviour. Many other types of behaviours can be compared with behavioural entropy.

Results

The goal was to assess whether the introduced behavioural entropy approach offers greater sensitivity when comparing driving data from healthy subjects and glaucoma patients. Analyses were performed on the 30m visibility conditions for all three speeds. Three metrics were calculated to facilitate the comparison: i) the InvTLC entropy introduced above (InvTLC Behavioural Entropy), ii) the standard deviation of lateral position (SDLP) and iii) the mean inverse time to line crossing (MeanInvTLC). These metrics are computed for each patient in each of the three speed segments and within the 30m visibility condition; this leading to 3 time 41 values for the 41 Glaucoma patients and 3 times 25 values of the 25 healthy subjects. The metrics are compared in terms of their ability to produce significant difference between healthy subjects and glaucoma patients in terms of Kolmogorov Smirnov statistic (KS-test) between each group's metric CDF (many other statistics could be used but again that is not the purpose of the paper). These CDFs do not take into account that some of the glaucoma patients have a more severe damage to the eye than others; this severity is measured by the mean deviation in a patient's ability to detect light dots in a dark field at many locations across their entire retina and is expressed in decibels (dB). A mean deviation of -10dB means that: the patient, on average, needed a 10 times higher intensity to detect the light dot than a normal person. A value of -20dB means a 100 times stronger intensity is needed for the patient to detect the light dot [11].

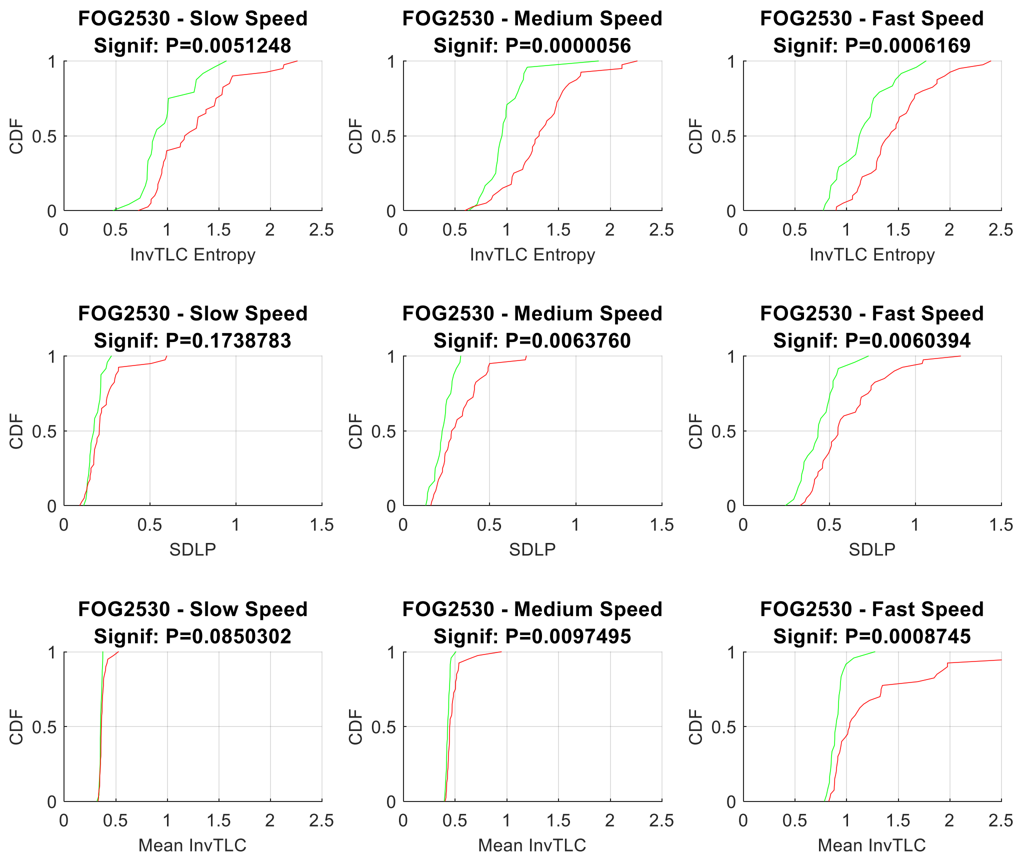


Figure 6. Cumulative distribution functions for healthy subjects (green) and glaucoma patients (red) for the three speed conditions in the 30m visibility condition (i.e. FOG2530).

KS-tests on the CDFs in Figure 6 that compare the two groups (healthy and glaucoma) shows that SDLP and MeanInvTLC do not yield a significant difference at the lowest speed but that InvTLC Behavioural Entropy does. In general the separation and therefore the significance between the two groups is much greater in the

Behavioural entropy case than in the other two cases indicating that Behavioural entropy offers superior sensitivity.

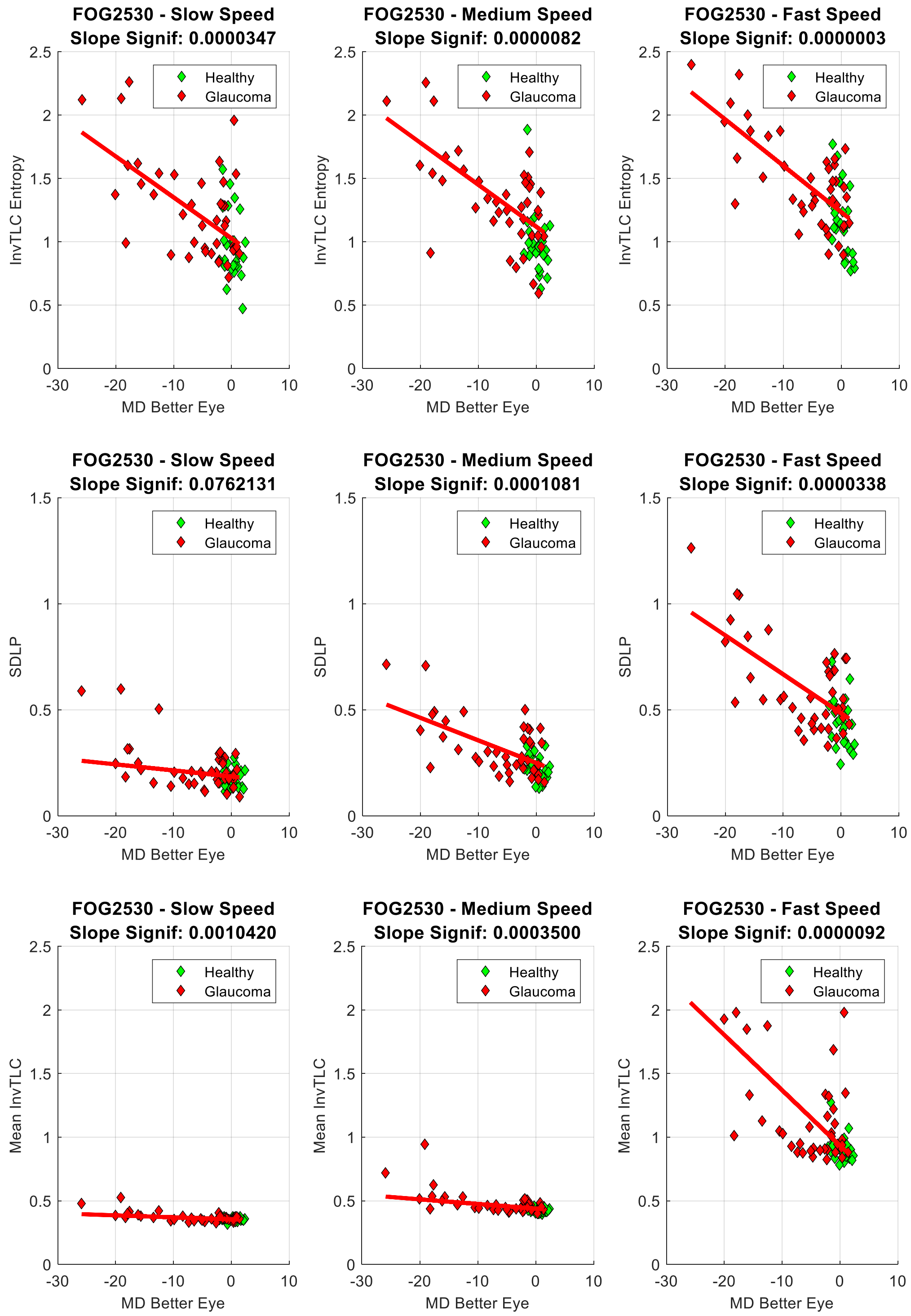


Figure 7. Effect of glaucoma severity (MD Better Eye) on curve negotiation in fog performance. Top panels show the effect in terms of entropy, middle panels in terms of SDLP and the bottom row of panels in terms of Mean InvTLC. The slope significance is the significance of the slope of the robust linear fit to the glaucoma patient performance as a function of disease severity expressed in mean deviation of the better eye (see text for explanation).

Understanding the relationship between driving performance and disease severity is crucial in the process of coaching patients about the elevated risks they may experience. The goal is to be able to sensitively determine what the incremental effect of disease severity (mean deviation of better eye) is on performance. In Figure 7 we show the robust regression lines together with their slope significance (Matlab R2015a RobustFit function [9]) that relate glaucoma patient's (red) performance as a function of disease severity for each of the three metrics in each of the three speeds. It is clear that InvTLC Behavioural Entropy produces a more sensitive assessment than the other two metrics based on the fact that the slope significances are greater in all three speeds than those observed in the SDLP and MeanInvTLC metrics.

Conclusions

The proposed behavioural entropy metric for assessment of differences in time series behaviours shows superior differentiation ability compared to more standard performance based metrics. It is important to note that behavioural entropy is not a performance metric per se but captures a shift towards more behaviours that are outlier (low probability) behaviours under the baseline condition. Behavioural entropy shows significances where more traditional metrics do not and shows levels of significance higher than that of traditional metrics. The first step in any behavioural analysis is to assess whether there is a detectable difference in behaviours across conditions. The behavioural entropy method appears to be a promising candidate. Once it has been established that a difference exist, exploration to determine whether the observed difference is of a meaningful magnitude commences. Here we advocate that a time based risk metric is most suitable because it quantifies how much time the driver (human operator in general) has before performance or safety is compromised.

References

1. Boer, E.R., Rakauskas, M.E., Ward, N.J., Goodrich, M.A. (2005). Steering Entropy Revisited. *3rd Intl. Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Maine (Jun 27-30).
2. Nakayama, O., Futami, T., Nakamura, T., and Boer, E., "Development of a Steering Entropy Method for Evaluating Driver Workload," *SAE Technical Paper* 1999-01-0892, 1999, doi:10.4271/1999-01-0892.
3. Goodrich, M.A., Stirling, W.C., Boer, E.R. (2000), "Satisficing Revisited". *Minds and Machines, Journal for Artificial Intelligence Philosophy and Cognitive Science*, Vol. 10, Issue 1, pp. 79-109, Feb.
4. Van Winsum, W., Godthelp, H., (1996). Speed choice and steering behavior in curve driving. *Human Factors*, 38(3), 434-441.
5. Godthelp, J., Milgram, J., & Blaauw, G.J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors*, 26, 257-268.
6. E.R. Boer, E.R., Ward, N.J. (2003). "Event-based driver performance assessment, " in *Proc. of the Second International Symposium on Driving Assessment, Training and Vehicle Design*, Park City, Utah, pp. 119–124.
7. Green, P. (2013). Standard definitions for driving measures and statistics: overview and status of recommended practice J2944. *In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '13)*. ACM, New York, NY, USA, 184-191.
8. McLean, J. R., Hoffmann, E. R. (1975). "Steering reversals as a measure of driver performance and steering task difficulty." *Human Factors*, 17, 248- 256.
9. Street, J.O., Carroll, R.J. Ruppert, D. (1988). "A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares." *The American Statistician*. Vol. 42, pp. 152–154.
10. Shabana N, Cornilleau Pérès V, Carkeet A, Chew PT. (2003). Motion perception in glaucoma patients: a review. Surv *Ophthalmol*. Jan-Feb;48(1):92-106.
11. Landers, J.; Sharma, A.; Goldberg, I.; Graham, S.L., (2010). "A comparison of visual field sensitivities between the Medmont automated perimeter and the Humphrey field analyzer". *Clinical & Experimental Ophthalmology*. February.
12. Shannon, C.E., Weaver. W. (1949). *The Mathematical Theory of Communication*. Univ of Illinois Press.

Computational Methods for Analysis of Visual Behavior Using Eye-Tracking

Alan Davies[1*], Andy Brown[2], Markel Vigo[1], Simon Harper[1], Caroline Jay[1]

[1*]alan.davies-2@manchester.ac.uk, [1]{fisrt.last}@manchester.ac.uk, [2]andy.brown01@bbc.co.uk
[1]School of Computer Science, University of Manchester, Manchester, UK
[2]BBC R&D, Salford Quays, Manchester, UK

Problem

Qualitative methods for analysing eye-tracking data are extreamly useful and widely used in the field of eye-tracking. These methods are however prone to incorrect assumptions [1]. We present a quantitaive method to determine the presence and extent of differences in visual behaviour between two groups, which is applied to the medical domain of electrocardiology. The electrocardiogram (ECG) is graph of the electrical activity of the heart and is a cheap and widely used medical test [2]. Misinterpretation of the ECG causes severe adverse events and even death in around 100,000 patients each year [3] and costs millions in litigation payouts [4]. Human over reading is still however considered the most accurate and relaible method of interpreting ECGs [5]. Interpretation of the ECG is predominatly a visual search task. This work aims to improve the understanding of gaze behaviour as a function of interpretation accuracy. Current methods of analysing eye tracking data are lacking or suboptimal in this area. Some work has been carried out examining the use of eye-tracking within the domain of ECG interpretation. Accuracy has been considered in terms of participants age and years of experience [6] and in the presence or absence of clinical history [7]. No studies to date have looked at accuracy from the perspective of eye movement behaviour in detail. One of the challenges to doing this is the lack of readily available tools and methods to accomplish such analysis. Some of the commonly used methods of aggregating visual behaviour, such as heat maps (figure 1) can provide qualitative information about the focus of attention and other metrics.

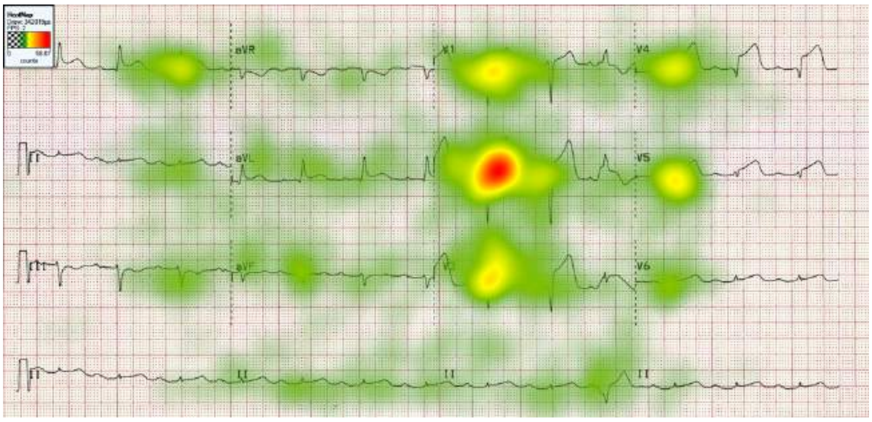


Figure 1: A 12-lead ECG with heat map overlaid. The redder the area the more visual attention received

These methods are less useful when a more quantitative approach to analysis is required. This work aims to provide quantitative alternatives to facilitate the comparrison of different groups visual behaviour to determine if eye gaze shifts are noticably different between medical staff making correct and incorrect ECG interpretations.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Methodology

Thirty one medical practitioners were shown a series of 12-lead ECGs on a computer screen that were taken from online open access ECG libraries. Each participant viewed eleven ECGs and spoke aloud their interpretation, which was recorded. Eye-tracking data was captured for each participant with a Tobii eye tracker. Areas of interest (AOIs) were mapped onto each lead of the ECGs using the Tobii studio eye-tracking software. An AOI is a region or area of a stimulus that researchers wish to collect data about [8]. For each stimulus the gaze data belonging to participants was split into two groups, those making correct interpretations and those making incorrect interpretations. Accuracy was based on the participant's interpretation matching the ECG descriptions from the source libraries. Common variations of the names of the described conditions would be accepted as correct. Insufficient detail, such as mentioning some of the salient features but not the name of the condition or stating a completely different condition would be categorized as an incorrect interpretation. The gaze data is then turned into a transition matrix for each group to reflect visual transitions also known as gaze shifts between the different AOIs. The transition matrices are then converted into 1st order Markov chains and the distance between the chains is measured using the Jensen-Shannon distance metric. The process is then repeated 10,000 times with a permutation test, with subsequent groupings being a random mixture of both correct and incorrect gaze data patterns of the same group sizes as the initial groups. These different distances can be plotted on a graph to generate a sampling distribution estimate. If visual transition is a factor of accuracy then we would expect to see a greater distance between the correct and incorrect groups than would occur by random chance. A p-value for this unknown test statistic can be expressed as a fraction of values at least as great as the non-permuted value. The method is summarized in figure 2.

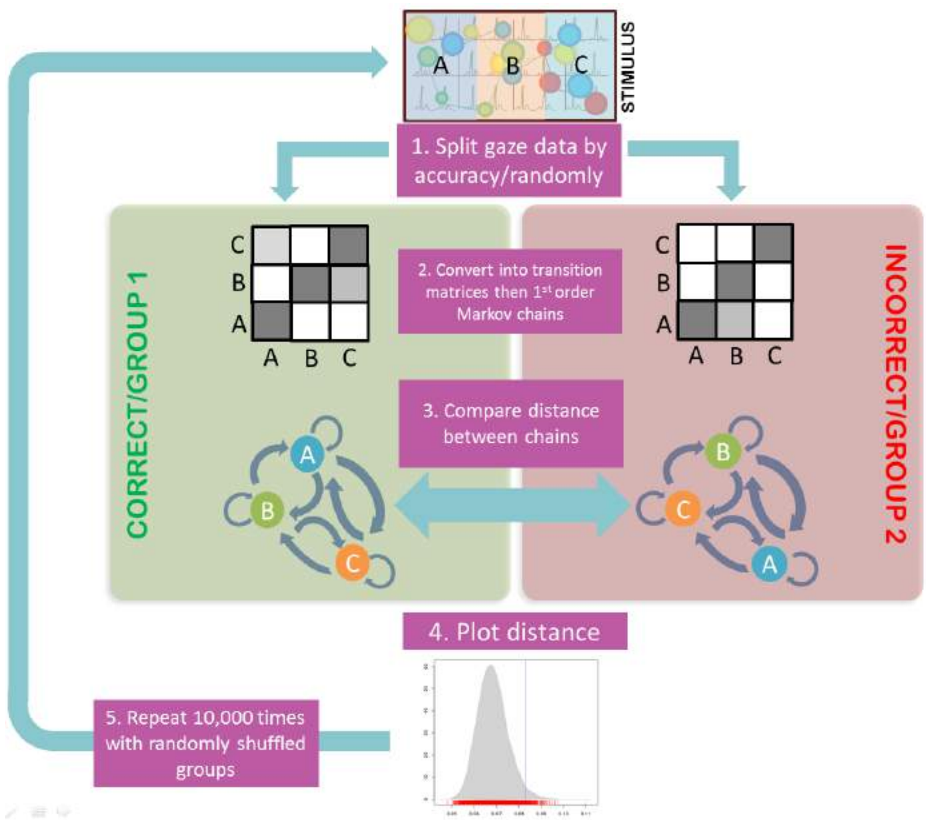


Figure 2: Illustrative overview of analysis method

This approach is used in order to overcome some of the challenges associated with highly variable, small sample sizes that are not amenable to more traditional statistical methods. Although the results are highly applicable to this particular sample, the technique described has the potential to be used as a method for exploratory data analysis with similar datasets.

Results to date

Using this method we found a statistically significant difference between the correct and incorrect groups in 5/11 stimuli ($\alpha \leq 0.05$). A summary of the results can be seen in table 1, which details the stimulus, distance, p-value and the sizes of the two groups.

Stimulus name	Jensen-Shannon distance	p-value	Group 1#	Group 2#
Anterolateral STEMI*	0.2884499	0.02	16	15
Atrial Flutter*	0.6041648	0.03	25	6
Hyperkalaemia*	0.6038147	0.03	2	29
LBBB	0.3634885	0.90	23	8
Normal Sinus Rhythm	0.4745519	0.20	23	8
Sinus Tachycardia	0.3218995	0.40	12	19
Supraventricular Tachycardia*	0.4400723	0.05	10	21
Torsades de points	0.4320926	1.00	5	26
Ventricular paced rhythm	0.3360781	0.70	9	22
Ventricular Tachycardia	0.5778351	0.50	26	5
Wolff-Parkinson-White syndrome*	0.3110491	0.04	12	19

Table 1: Results per stimulus (* = $p \leq 0.05$)

Future work

Generation of top-down AOIs can introduce bias and relies on certain assumptions made by researchers. We propose an iterative grid based system that can be used to help determine the best level of granularity for a particular stimulus by systematically varying the AOI size. In the case of the ECG it may be a much smaller unit than the lead. This would seem probable as we know practitioners are trained to look at specific parts of the ECG waveform pattern. This will help us to identify at which resolution the difference between two groups is at its greatest.

Conclusion

This method allows us to ask questions about the differences between aspects of visual behaviour that occur between two different groups. It also reduces the visual complexity of some of the standard eye-tracking visualizations whilst providing a quantifiable measure of difference. The benefit of the permutation tests also allow for analysis of small groups that may be significantly different in size.

References

1. A. Bojko, *Eye tracking the user experience: A Practicle Guide to Research*. New York: Rosenfeld Media, 2013.
2. A. Davies and A. Scott, *Starting to read ECGs: A Comprehensive Guide to Theory and Practice*. London: Springer-Verlag, 2015.
3. H. Holst, M. Ohlsson, C. Peterson, and L. Edenbrandt, “A confident decision support system for interpreting electrocardiograms,” *Clin. Physiol.*, vol. 19, no. 5, pp. 410–418, 1999.
4. P. Mele, “Improving electrocardiogram interpretation in the clinical setting.,” *J. Electrocardiol.*, vol. 41, no. 5, pp. 438–9, 2008.
5. A. E. Buxton, H. L. Kennedy, B. P. Knight, J. a Y. W. Mason, C. D. Schuger, W. L. Winters, and A. W. Boone, “ACC / AHA CLINICAL COMPETENCE STATEMENT ACC / AHA Clinical Competence

Statement on Electrocardiography and Ambulatory Electrocardiography A Report of the ACC / AHA / ACP – ASIM Task Force on Clinical COMMITTEE MEMBERS TASK FORCE MEMBERS TABLE OF CONTENTS," *J. Am. Coll. Cardiol.*, vol. 38, 2001.

6. R. R. Bond, T. Zhu, D. D. Finlay, B. Drew, P. D. Kligfield, D. Guldenring, C. Breen, a G. Gallagher, M. J. Daly, and G. D. Clifford, "Assessing computerized eye tracking technology for gaining insight into expert interpretation of the 12-lead electrocardiogram: an objective quantitative approach.," *J. Electrocardiol.*, vol. 47, no. 6, pp. 895–906, 2014.
7. G. Wood, J. Batt, A. Appelboam, A. Harris, and M. R. Wilson, "Exploring the Impact of Expertise, Clinical History, and Visual Search on Electrocardiogram Interpretation," *Med. Decis. Mak.*, vol. 34, no. 1, pp. 75–83, Jan. 2014.
8. K. Holmqvist, M. Nystrom, R. Anderson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. New York: Oxford University Press, 2011.

Measuring Brain based Specific Learning Difficulties: Subtypes of ADHD and Dyslexia distinguished by Electrophysiology (Quantitative EEG and Event Related potentials)

T. Steffert[1] and B. Steffert[2],

[1]**Centre for Research in Computing, The Open University, Milton Keynes, England. b.steffert@mail.bbk.ac.uk**
[2]**Learning Recovery, Cambridge, England.**

Abnormalities in quantitative Electroencephalogram (qEEG) and Event Related potentials, (ERPs) in ADHD, ADD and Dyslexic patients have been reported for some decades.

This talk will describe the different patterns of brain activity measured by the combination of quantitative EEGs and ERPs and how they can help to reveal more specific subtypes of a range of conditions and with additional psychometric and genetic testing, how these measurements can form "endophenotypes" [1] which can lead to better tailored treatments and improved prognoses.

Quantitative EEGs

A quantitative EEG records oscillations in the electrical activity of the brain. Variations in alertness and behavioural control are directly related to specific EEG rhythms that emerge over cortical areas of the brain as it is activated during specific tasks from sleeping, dreaming, thinking, feeling, hearing, seeing or moving. Different functional areas resonate differently according to the activity but in ADHD and Dyslexia there are patterns consistently associated with difficulty in activating the necessary frequencies. For example many ADHD individuals show excess low frequency (the left side of the arousal curve in Figure 1) throughout their frontal lobes which reflects their difficulty in initiating and maintaining attention on tasks of minimal interest to them, such as schoolwork in children or paperwork in adults. On the other hand a brain set too high (the right side of the arousal curve) will be constitutionally anxious.

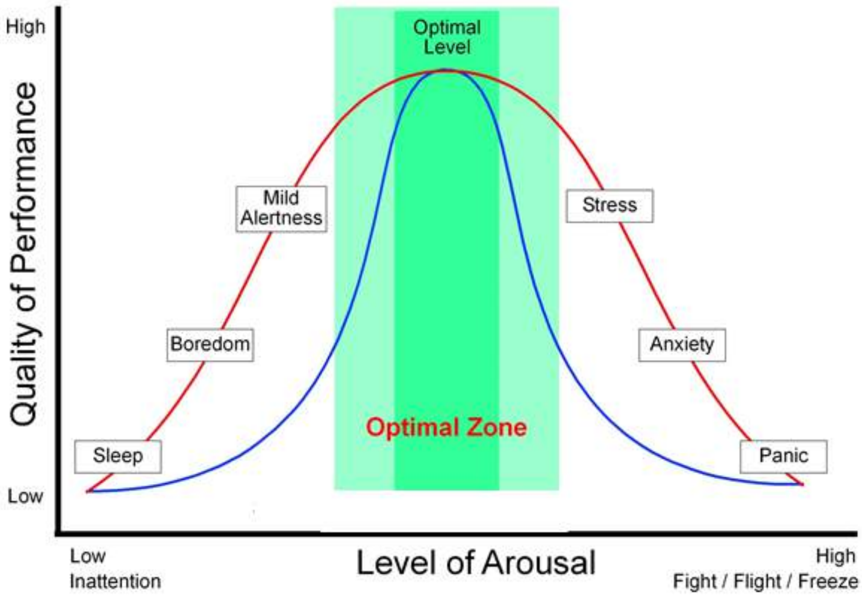


Figure 1: Schematic adapted from: The Yerkes–Dodson [2] arousal and performance curve. The red line is an example of a person who would have a better performance over a greater range of arousal and a wider optimal zone of arousal (in green) in comparison to the blue line.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Different functional areas of the brain resonate differentially, in terms of speed and power according to the task, and can be measured with electrodes from the scalp. Of all brain imaging techniques, EEG/ERP provides a better temporal resolution of brain activity [3] and is a functional measure of brain-based problems, as opposed to a structural one, such as an MRI.

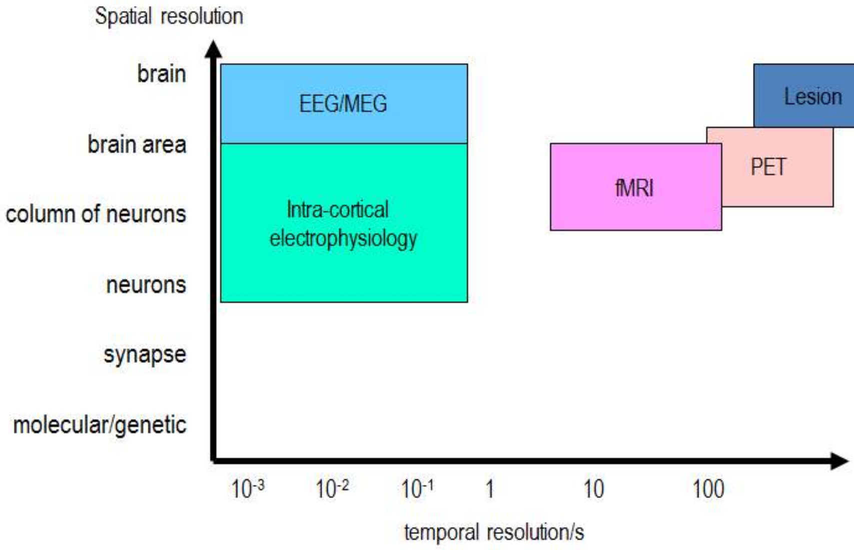


Figure 2: Schematic adapted from [4]: The horizontal axis shows time, ranging from milliseconds to hundreds of seconds and the vertical axis shows spatial resolution.

A qEEG is measured from 19 or more functional areas of the scalp and the names of the locations are given by the International 10-20 electrodes placement system [5].

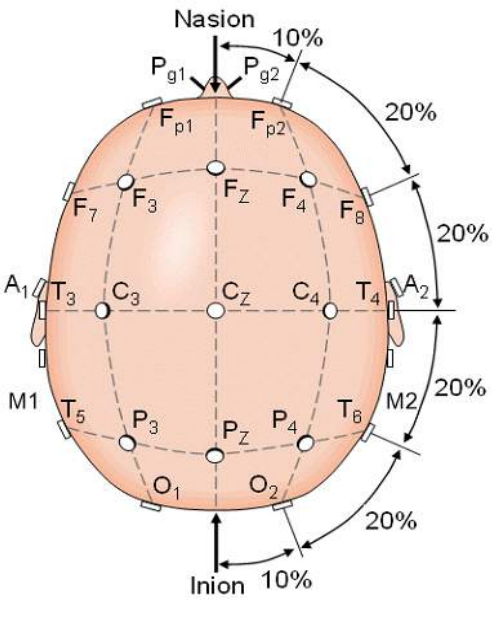


Figure 3: Schematic of the International 10-20 electrodes placement system. Where Fp stands for Frontal Pole, F stands for Frontal, C is for Central, T is of Temporal, P is for Parietal and O is for Occipital. The odd numbers are on the left and even on the right, and z is for the centre line and the numbers increase as they move out from the centre line.

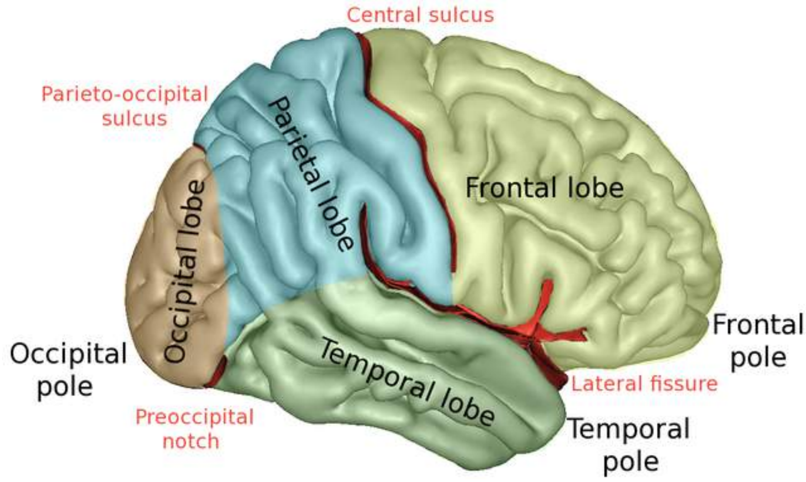


Figure 4: Schematic of Brain lobes, [6]

The electrical activity of the brain is picked up by electrodes and displayed to a computer (Figure 5). Functional areas are the lobe of the brain over which the electrode is placed (Figure 4). For example a simple overview of the research with particular regard to school problems shows us that the Executive system of the brain is based on the Frontal lobe. Dysfunction here is associated with impulsivity, disorganization, difficulty in planning, time keeping, maintaining attention and various other Executive dysfunctions. The temporal lobes mediate memory and house the auditory cortex while the occipital lobes manage visual information so dysfunction here can cause literacy problems among other things. The parietal lobes specialize in shape and space perception and lack of activity here can mean difficulties in catching balls, handwriting .and perception of non-verbal communication. Each area has a network that connects to other functional areas when necessary, which when disturbed gives an arrhythmia leading to inefficiency in performance of the activity.

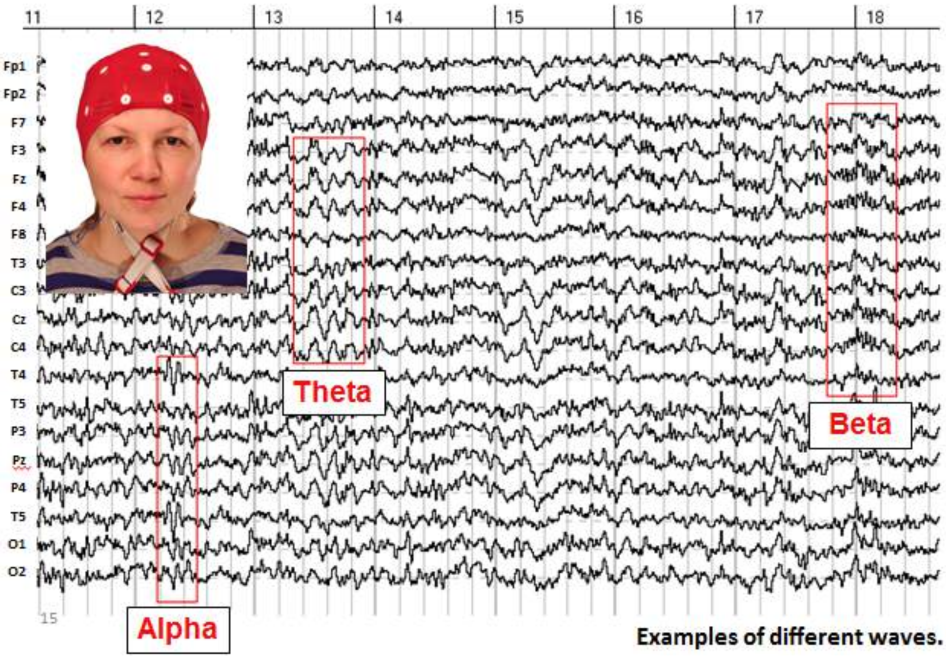


Figure 5: Example of a 19 channel EEG and electrode cap. The red boxes highlight the different frequencies of activity seen in the raw EEG.

Attention Deficit Hyperactivity Disorder

Sub-types of ADHD can be identified by qEEG. For example, the major pattern associated with the diagnosis of ADHD, impulsivity, inattention and hyperactivity, (representing around 60% of the ADHD population) is shown by an excessive low frequency (4–7 Hz) that can be combined with a deficit of high frequency (18-21 Hz) EEG activity (Figure 6).

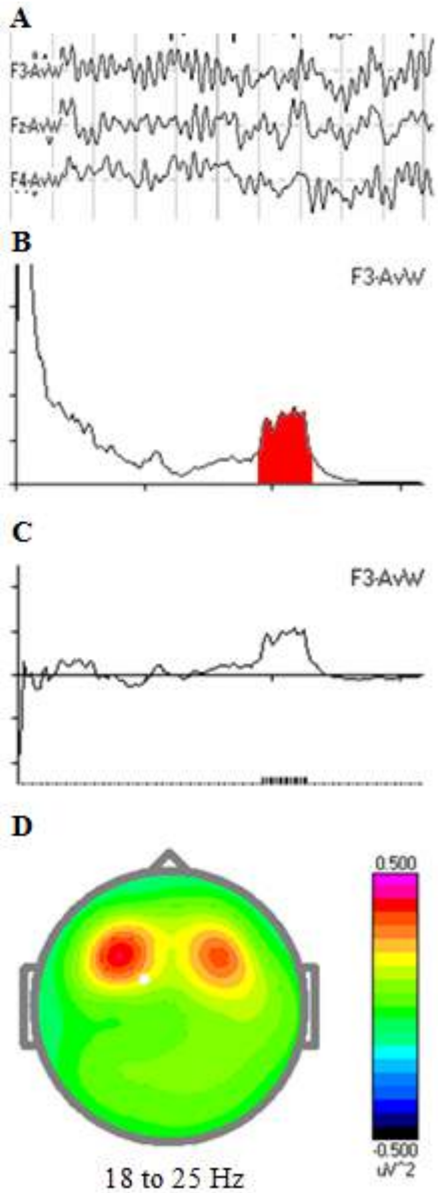


Figure 6: Shows excessive low frequency activity in the frontal and central lobes that is a marker of hypoarousal seen in a major subgroup of ADHD. Fig 6A, shows 2 seconds of raw EEG from three frontal and central electrodes, clearly dominated by 7hz activity. Fig 6B, shows the average spectra from 1 to 30 Hz of 3 min of EEG from the left frontal cortex Cz, where the red peak shows the 7Hz activity. Fig 6C, shows the statistical comparison of the spectra with a normative database, were the trace goes up from the centre line this indicates more activity in the person in comparison to healthy age matched controls and vice versa, the small black squares below indicate the statistically excess activity in the low frequency range. Fig 6D, shows a topological head map of the statistically excess activity at Cz.

A second typology, the “inattentive only” subtype is characterised by excessive middle range frequencies of 7-12 Hz known as Alpha rhythms. This is the dreamy child.

A smaller proportion (around 5 to 7%) have the opposite pattern, of 'hyperarousal' with a raised amplitude of high frequency brain activity (above 21 Hz) (Figure 7). This underlies the irritability and aggressiveness of some clients and for which stimulant medication would be contraindicated.

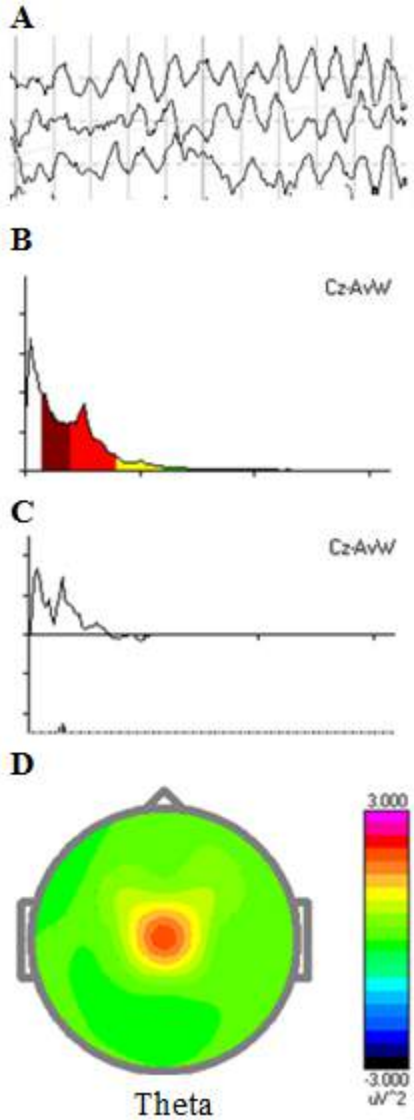


Figure 7: Shows excessive beta activity in the frontal lobes that is a marker of hyperarousal seen in a small subgroup of ADHD. Fig 7A shows 2 seconds of raw EEG from three frontal electrodes. Fig 7B shows the average spectra from 1 to 30 Hz of 3 min of EEG from the left frontal cortex F3, the red shaded area is the excess beta activity. Fig 7C shows the statistical comparison of the spectra with a normative database, the small black squares below indicate the statistically excess activity in the 18 to 25 Hz range. Fig 7D shows a topological head map where red indicates the excess activity.

For example the well know theta/beta ratio that is associated with ADHD, and which may actually be only a maturation lag [7] is shown in (Figure 8).

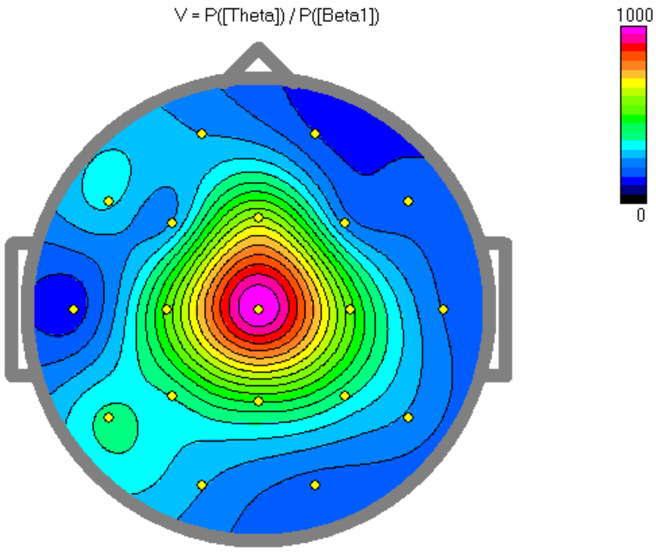


Figure 8: Shows the topological distribution of the Theta to Beta ratio which here is 10/1 ratio at Cz, i.e. 10 time as much Theta (the under arousal brainwaves) than Beta (the attention brain waves). The Theta/Beta ratio is usually measured at Cz, in the centre of the cortex and is typical not more than 2 to 1 in adults and 4 to 1 in children.

Event Related Potentials

Event-Related Potentials (ERPs) are a further measure derived from the EEG and represent the electrical brain potentials prompted by cognitive processing of a stimulus and reflect internal (endogenous) mental processes. The test/retest reliable of ERPs is better than EEG measures. So, for example, a well-known ERP is called the Mismatch Negativity (MMN). This is a negative shift in electrical brain potentials in response to an odd or rare stimulus in a long sequence of 'typical' stimuli and is therefore not dependent on the nature of the stimuli per se, but it's meaning or higher order properties, such as novelty – the unexpected.

ERPs are elicited by the presentation of hundreds of short stimuli. These stimuli are then averaged to cancel out any random background activity that is not synchronised with the onset of the stimuli or the response of the person. ERP's have both positive and negative voltage fluctuations in comparison to a baseline directly preceding the stimulus presentation. These fluctuations are characterised by the timing and amplitude of the peaks and troughs of their activity, as well as the scalp topology. So, for example, a positive deflection at 300 milliseconds is called the P300 and a negative-going trough at around 100 milliseconds after the presentation of the stimulus is termed the N100 or N1 for short.

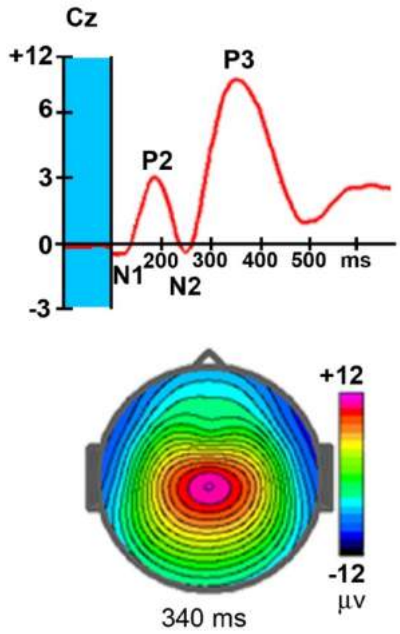


Figure 9: Schematic of a stereotypical Evoked and Event-Related response, with time along the x axis and amplitude along the y axis. The blue shaded area is the 100 ms when the stimulus was presented and the red line shows the positive and negative voltage fluctuations at different time points, elicited by the stimulus and the cognitive processing of the stimulus. The lower half shows the topological scalp distribution of the peak amplitude at 340 ms after the stimulus onset.

Event-Related Potentials reflect the time course of sensory and cognitive processing and can help to distinguish between sensory and cognitive deficits even when there is no external behavioural response such as a button push. ERPs have excellent temporal resolution, as quick as 1 millisecond, but the amplitude is very small, so they require hundreds of trials to be averaged. Therefore, a typical task can take around 20 minutes.

ERP's then, represent a measure of the brain based reaction to sensory perception and cortical response which is prior to other physiological measures such as facial, emotional, physical and cognitive responses. They offer a neurobiological marker that is more objective than psychometric tests and questionnaires.

In the ERP analysis of ADHD one of the most consistent findings is the ability to inhibit a prepared response and is highly associated with an executive function that predicts impulsivity,

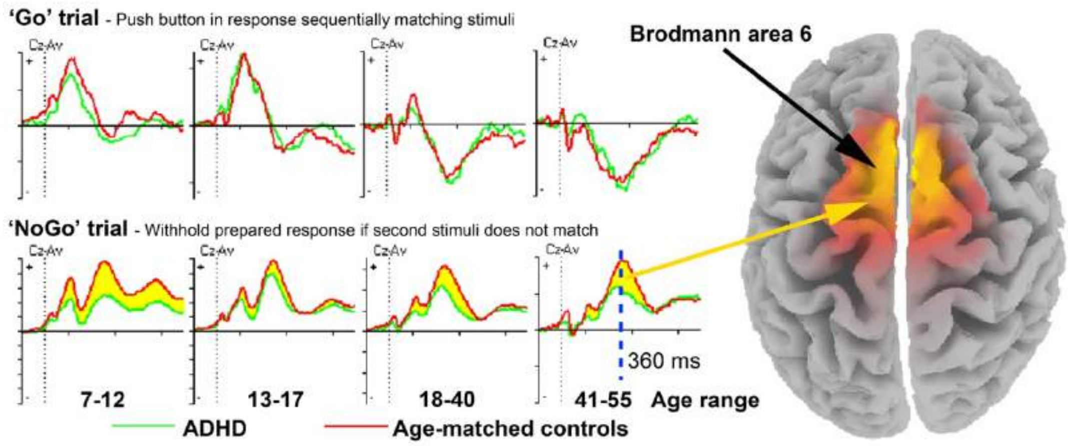


Figure 10: Reduction of the action inhibition component in ADHD subjects.

The Go/NoGo task is to push a button as quickly as possible when an image of an animal is followed by a second animal 1100 milliseconds later – this is the 'Go' trial. The subject is instructed to withhold the planned motor response if the second image is not an animal – this is the 'NoGo' trial. The top row of four event-related potentials (ERPs) in the different age ranges are elicited in the 'Go' trial and the bottom row shows ERPs from the 'NoGo' trial, with ADHD subjects shown in green and age-matched controls in red. The yellow shaded area is the difference between ADHD subjects and controls. It shows no significant difference in the 'motor engagement component' in the top row, but demonstrates a significant difference across all ages in the 'P300 action inhibition NoGo' component on the bottom row; that is, impulsive responding. The image of the brain on the right shows the source of this activity by sLORETA [8] to be the Brodmann area 6, the supplementary motor area of the frontal lobe

Other functional brain markers that predict unstable emotional, cognitive and physical states that leave individuals unable to draw on normal inhibitory and excitatory mechanisms are measured by Event Related potentials. ERP's can measure Executive Function, Working Memory, Speed of Processing, impulsivity, attention, and emotional control.

Dyslexia

Dyslexia is generally defined as a discrepancy between one or more literacy skills on the one hand and cognitive, numerical and/or other academic skills on the other; that is not due to lack of educational opportunity or socio/emotional hindrances. There are of course degrees and differences in the exact type of reading difficulty which are now better encapsulated by the new DSM-5. [9]

The strongest correlate of all proposed causes of Dyslexia is genetic and there are several genes identified which are involved in neuronal migration and axonal growth which affect neuronal membrane efficiency [10]. Prior research [11] had already shown that the integrity and timing of the connections between functional areas of the reading circuit in the brain, differed between poor and fluent readers and that the decoding of a word requires breaking down the sound sequence of the word into phonemes, from which all words are composed, whilst simultaneously ordering the graphemes (letters and letter clusters) from left to right in exact orthographic sequence.

Goswami, [12] looked at the effect of this auditory deficit on word structure, finding that the elements of a word, such as the rise time of the first sound and the perception of the stressed syllable in the word (which is denoted as the amplitude envelope) were the critical hurdles to efficient reading for Dyslexics. Thus both timing and amplitude are correlates of efficient reading and the difficulty in processing amplitude onsets accurately, in Goswami's view may constitute the primary deficit in developmental Dyslexia.

But evidence for visual processing problems has existed from the first identification of "word blindness" in 1909. The visual and auditory systems that are specialised to respond to temporal transients run from the retina to the visual and temporal cortex, these are the magnocellular pathways which feed into the dorsal stream, controlling visual and auditory attention, eye and articulatory movements.

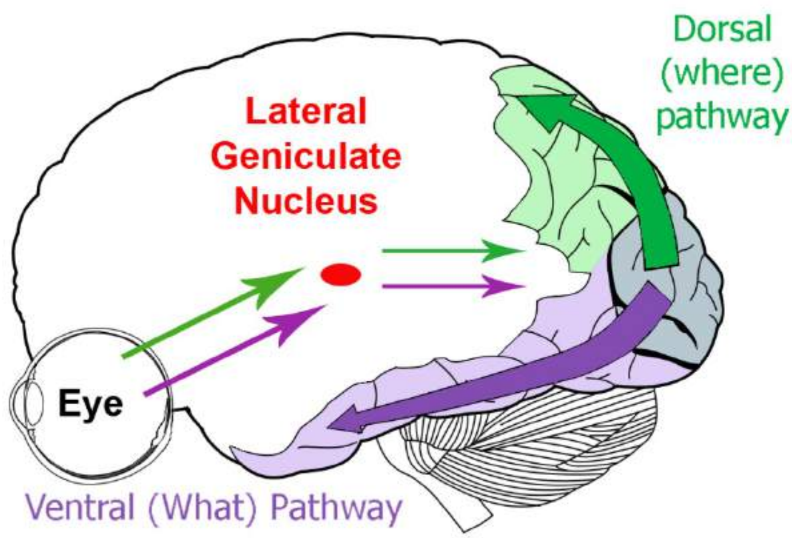


Figure 11: Magnocellular and parvocellular are two separate visual pathways going from the eyes to the Lateral Geniculate Nucleus (LGN) of the thalamus and on to the visual cortex. The magnocellular pathway from LGN feed into the dorsal visual stream in the parietal cortices and resolving motion and coarse outlines detection and estimate where things are in space. Whereas the parvocellular pathway goes to the ventral visual stream and has a higher spatial resolution but lower temporal resolution and is also sensitive to colour.

When the magnocellular pathway is compromised by patchy myelination as [13, 14] has shown, there is a slower, more inaccurate, visual processing which causes unfocussed visual attention & unstable eye control as well as letter order confusion since tiny details of the visual pattern of print have to be accurately segmented into separate letters in order to match them with the phonemes they represent, mistakes are more likely.

Perceiving fine visual detail is not necessary for most of the visual world, but tiny details of letters have to be accurately sequenced in order to match them to the phonemes they stand for. So inefficient visual sequencing means the reader cannot track rapid changes in visual signals and letters can appear blurry or to move because the Dyslexic cannot achieve stable binocular control.

Our study [15] showed a deficit in the P300 no go component - a dorsal visual component and indicating a deficit in visual rise - time, shown in the ERP's which seems to correlate with the rise time in the onset of speech and also the stressed syllable, shown by Goswami [12] to be important predictors of reading difficulties. The rise time in reading is an acoustic property while rise time in ERP's reflects the rate at which sensory information enters the brain - but these seem to be correlated. Whether one predicts the other or whether there is an underlying factor predicting both slow auditory and visual rise time remains to be determined.

Our subjects from a clinical sample gave informed consent and ranged from 7 years to 40 years which suggests that the dorsal pathway deficit identified is not a consequence of reading, it is pre -existing, heritable and long lasting.

Figure 12 shows the age range of Dyslexics (7 years to 40 years) all showing the same deficit, suggesting that this slowness exists pre-reading and other research suggests is hereditable.

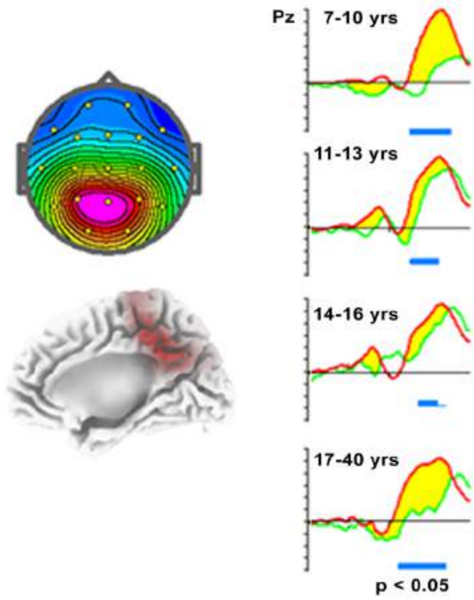


Figure 12: The Independent component (IC) of the P3 b posterior of the GO trial, in the Visual Continuous Performance Task (VCPT), Event-Related Potentials. On the right: the green line shows the Independent Component of the ERP of the dyslexics and the red line of the age matched healthy control at Pz. The yellow shaded area shows the difference between the two and the blue horizontal lines indicate the time interval with statistically significant ($p < 0.05$ by t-test) differences between the dyslexics and healthy control groups. On the *Y*-axis: averaged potential, each mark corresponds to 2 μV. The *X*-axis is time after the onset of the second GO stimulus. The whole time interval corresponds to 400 ms. Numbers (such 7-10, 11-13 etc.) indicate age range for a group. On the left – topography and sLORETA image of the independent component. The independent component is back projected to the electrodes. There was a statistically significant difference between the dyslexics and corresponding healthy age matched control groups, in the Independent component of the ERP generated in the parietal cortex (Figure 12). This component appears to be associated with activation of the dorsal visual stream, known as the P3 GO posterior. It could be also considered as the visual aspect of the conventional P3b (or target P3) wave.

The conclusion from our study is that Dyslexia is characterised by slow visual and/or auditory sensory processing which affects the acquisition of the phonological and orthographic codes necessary for reading and can therefore be best subtyped through ERP's, which measure the speed of processing through the brain.

General Conclusion of measuring brain based behaviour by qEEG and ERPs.

Thus the combination of both quantitative Electroencephalogram (qEEG) and event related potentials (ERP's) offer a more specific analysis than psychological or physical measures alone. When added to psychometric measures, these biological measures can offer the basis of an endophenotype that forms a link between the genotype and phenotype. This allows more accurate diagnosis and prognosis. Medication for example is more efficacious for some subtypes of ADHD than others and recent research suggests that this can be tested within 20 minutes by recording a pre and post qEEG/ERP after a dose of a stimulant such as Ritalin. [16, 17]

Furthermore treatments such as Neurofeedback or Transcranial Direct Current stimulation that can help remediate these conditions is more effective when subtypes are identified since more accurate placements of electrodes that can alter brainwaves are indicated. Even psychological treatments such as Cognitive Behavioural Therapy can benefit from knowing whether for example the inattention of a student is due to their excessive low frequency brainwaves (meaning mental energy is difficult to summon up) or from crippling anxiety generated by excessive high frequency brain waves. Treatment also differs in Dyslexia with the visual pathways being better remediated by visual exercises, spectrum specific lenses or prisms, while auditory deficits benefit more from phonological training. Subtyping most psychological problems such as Post traumatic Stress, depression etc. will yield similar benefits in terms of understanding and outcomes.

Additionally, the EEG recording equipment is becoming cheaper, smaller and easier to use [18], with automated analysis software, producing Real-Time metrics measuring an array of states from attention to relaxation, allowing this powerful tool to escape from the laboratory and the clinic, on to the sports field [19] or boardroom. Applications are being explored for example in non-pathological conditions such as creativity, elite performers and healthy ageing and normative databases of elite performers are being collected.

References

1. Ritsner, M. S. (Ed.). (2009).The Handbook of Neuropsychiatric Biomarkers, Endophenotypes and Genes: Volume I: Neuropsychological Endophenotypes and Biomarkers (Vol. 1). Springer Science & Business Media.
2. Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology **18**, 459-482.
3. Kropotov, J. D. (2010). Quantitative EEG, event-related potentials and neurotherapy. Academic Press.
4. Meyer-Lindenberg, A. (2010). From maps to mechanisms through neuroimaging of schizophrenia. Nature **468**, 194-202.
5. Jasper, H. (1958). Report of the committee on methods of clinical examination in electroencephalography. Electroencephalogr Clin Neurophysiol **10**, 370-375.
6. By Sebastian023, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=21020857
7. Arns, M., Conners, C. K., & Kraemer, H. C. (2012). A decade of EEG theta/beta ratio research in ADHD: a meta-analysis. Journal of attention disorders, 1087054712460087.
8. Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp Clin Pharmacol **24**, 5-12.
9. DSM-5 American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders. Arlington: American Psychiatric Publishing.
10. Shaywitz, S. E., Shaywitz, B. A., Pugh, K. R., Fulbright, R. K., Constable, R. T., Mencl, W. E., & Katz, L. (1998). Functional disruption in the organization of the brain for reading in dyslexia. Proceedings of the National Academy of Sciences **95**, 2636-2641.
11. Spironelli, C., Penolazzi, B., & Angrilli, A. (2008). Dysfunctional hemispheric asymmetry of theta and beta EEG activity during linguistic tasks in developmental dyslexia. Biological Psychology **77**, 123-131.
12. Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. Reading and Writing **23**, 995-1019.
13. Klingberg, T., Hedehus, M., Temple, E., Salz, T., Gabrieli, J. D., Moseley, M. E., & Poldrack, R. A. (2000). Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. Neuron **25**, 493-500.
14. Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. Trends in neurosciences **20**, 147-152.
15. Steffert and Steffert (2014) Deficits in the Dorsal Component of the P3b Executive ERP in Dyslexia. Conference proceedings Creating Impact through Innovation p140 The British Dyslexia Association Conference 2013.
16. Arns, M., Gunkelman, J., Breteler, M., & Spronk, D. (2008). EEG phenotypes predict treatment outcome to stimulants in children with ADHD. Journal of Integrative Neuroscience **7**, 421-438.
17. Ogrim, G., Kropotov, J., Brunner, J. F., Candrian, G., Sandvik, L., & Hestad, K. A. (2014). Predicting the clinical outcome of stimulant medication in pediatric attention-deficit/hyperactivity disorder: data from quantitative electroencephalography, event-related potentials, and a go/no-go test. Neuropsychiatric disease and treatment **10**, 231.
18. Väljamäe, A., Steffert, T., Holland, S., Marimon, X., Benitez, R., Mealla, S., ... & Jordà, S. (2013). A review of real-time EEG sonification research. Presented at the 19th International Conference on Auditory Display (ICAD2013) on July 6-9, 2013 in Lodz, Poland.
19. Thompson, T., Steffert, T., Ros, T., Leach, J., & Gruzelier, J. (2008). EEG applications for sport and performance. Methods **45**, 279-288.

The Importance of Behavioural Bioassays in Neuroscience

R.E. Brown and S. Bolivar

Department of Psychology and Neuroscience, Dalhousie University, Halifax, Canada; rebrown@dal.ca

Abstract

The behavioural bioassay is fundamental to research in behavioural neuroscience. A described by Tinbergen, behaviour is measured to answer questions about development, mechanisms, adaptation and evolution. Chemical assays, bioassays, and behavioural bioassays have been developed for detecting and quantifying substances such as neurotransmitters, hormones, and toxins and for measuring behaviour. This paper begins with an overview of these methods and then focuses on how behavioural bioassays are developed. Ethograms and qualitative descriptions of behaviour units are discussed. Sampling and recording rules are then considered, along with quantitative descriptions of the behaviours being observed. The paper concludes with examples of behavioural bioassays used for detecting various internal and external stimuli, along with considerations such as the complexity of the stimuli and the problem of measuring "psychological" states such as anxiety, from behaviour. Suggestions are made for improving the validity and reliability of behavioural bioassays in neuroscience.

1. Why measure behaviour?

We measure behaviour to answer a question about an animal or a person. Tinbergen [1] showed that questions about behaviour could be divided into four general categories: ontogeny, mechanism, adaptive significance and evolution [2] [3]. Thus behaviour is used to answer questions posed in terms of experiments. For example, the ontogeny question: "Are there age-related changes in vision in mice of strain X?" could be answered by developing a mouse vision test apparatus and procedure, and then testing mice of different ages in this apparatus [4]. The behaviour of the mice is then used to answer the question. A mechanistic question might be: "Does damage to the hippocampus disrupt learning and memory in mice?" Again, an apparatus and procedure for testing memory would be developed, a lesion made in the hippocampus of the mice, and the behaviour of these mice in this test used to answer the question. In each case, a behavioural bioassay has been developed to answer the experimental question. In a sense, all studies of animal behaviour [5] are behavioural bioassays.

2. Three types of assays.

An assay is a method used in chemistry, biology, and neuroscience to qualitatively assess the presence of a particular substance (X) within a sample, and/or quantitatively measure the amount of that substance within a sample. Assays are commonly used in neuroscience to measure hormone levels and neurotransmitter levels or to detect the presence of toxins in the body. Different assay methods such as chemical assays, bioassays, and behavioural bioassays have been developed. The purpose of this paper is to conceptualize behavioural bioassays in the context of chemical assays and bioassays and to discuss the problems of reliability and validity of these assays.

2.1. Assay methods in chemistry.

In chemistry, assays are used for quantitative or qualitative evaluation of a sample for a particular substance (X). For example, a rock could be assayed to qualitatively detect the presence of gold and to quantitatively measure the amount of gold present. This process requires that the gold be dissociated from the rest of the rock's components using a number of different assays, such as the nickel sulphide fire assay coupled with flame atomic absorption spectrometry [6] or inductively coupled mass spectrometry [7].These methods of analysis require that a series of well-defined steps be performed in a particular sequence taking care not to make any omissions or substitutions. Having a standardized procedure is essential because the accuracy of a method can be affected by

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

many factors including the chemicals used during the dissociation process and the purity of the sample in which the gold is found. When the concentration of the gold is determined through a calibration curve, the accuracy of this curve will dictate the precision of the gold measurement. Repeated measurements of the same sample establish the reliability of the measure used, while comparisons with a known concentration of gold in a sample will establish the validity of the measurement. Once a valid and reliable assay procedure has been established, each step in this protocol is necessary and must be performed in the correct order for accurate detection, extraction, and measurement of the gold.

2.2. Bioassays

A bioassay uses a biological response (Biosensor) to detect the presence of a particular substance (X) and quantify the amount of that substance in a sample [8]. The biosensor can be a whole plant or animal, vertebrate or invertebrate (in vivo bioassay), an isolated organ (ex vivo bioassay), a cell culture (in vitro cellular bioassay) or an isolated sub-cellular system (in vitro enzyme assay, immunoassay, receptor assay, etc.). In order to respond to X, the bioassay must have a sensor for X and produce a quantifiable biological response to X. The greater the response, the greater the amount of X in the sample. This response can be standardized and then a standard biological response curve can be calculated. Bioassays can produce a graded dose- response curve or a zero-one response in which an all-or-none reaction occurs, such as cell death, when the substance exceeds a threshold amount. Bioassays can be used to detect the presence of substances in the external environment such as chemicals, toxins, and environmental pollutants and to detect internally generated substances such as hormones, enzymes and neurotransmitters. As with chemical assays, bioassays have a series of steps that must be conducted in a specific order with no omissions or substitutions. For example, when performing a radioimmunoassay (RIA) of testosterone a procedure involving 25 steps has been developed [9]. As with the gold assay, it is important that all steps in this process are carefully carried out to completion to obtain an accurate analysis of the testosterone that is present.

2.2.1. Cell culture bioassays

Cell-based bioassays include assays which measure the number of live cells (viability assay), the number of dead cells (cytotoxicity assay) or the mechanism of cell death (eg., apoptosis) and are often used for screening new drugs [10]. Cell culture bioassays can also be used for determining the pharmacological activity of novel substances, the activity of endogenous proteins, drug toxicity, and measuring pollutants or toxic chemicals in the environment. An important consideration when using bioassays is whether or not the tissue that is used to sense the presence of the substance (X) is selective for that substance, or can be confounded by responses to other substances similar to X. Additionally, when using whole organisms many possible confounds should be considered such as the age, genetic background, and sex of the animal. As with the assay for gold, the validity and reliability of bioassays must be determined. For example, a number of different cell culture cytotoxicity bioassays can be used to measure the toxicity of arsenic trioxide [11]. The instructions for using the SH-SY5Y cell culture bioassay are quite specific [12] and failure to follow the procedure exactly as outlined can lead to problems with the validity and reliability of the assay. The MTT cell proliferation assay for SH-SY5Y cells has a complete set of instructions (available on line) which has a section on troubleshooting. It is clear from this example that there are many ways for cell culture bioassays to go wrong.

2.2.2. Antibodies

Many bioassays depend on antibodies that selectively bind to a molecule (X) that is to be detected and quantified in the sample. If the antibodies have batch-to-batch variability or bind to molecules other than the one that they are designed to detect (X), they lack reliability and validity [13]. Poorly characterized antibodies thus provide one of the most significant problems in the use of bioassays [14] [15]. The quality of the antibodies used is thus a key problem in the reliability and validity of many bioassays, and efforts are being made to develop standards for antibodies in order to increase the reliability and validity of bioassays [16] [17].

2.2.3. Other sources of error in using bioassays

As shown by Plebani [18], there are many sources of error in using bioassays, including pre-analytical errors, equipment malfunction, analytical and post-analytical errors. Latendresse and Ruiz [19] give examples of the types of errors made in conducting bioassays for CRH levels in the blood. There are many suggestions as to how such errors can be prevented to increase the reliability and validity of bioassays [19] [20].

2.2.4. Assays are only as good as their reliability and validity

It is often argued that behavioural testing is difficult and subject to errors. The preceding discussion on chemical assays and bioassays indicates that these highly developed assays have many problems with reliability and validity and that constant vigilance is required in order to perform the assays properly. Bioassays are complex; standard procedures are not always followed; there are many possible sources of error; the assays must be constantly recalibrated for reliability and validity; and simply because an antibody or a "kit" is purchased from a reliable supplier, there is no guarantee that the antibodies in the kit do what they are supposed to (buyer beware).

2.3 Behavioural Bioassays

Behavioural bioassays measure an organism's behaviour, qualitatively or quantitatively, to detect and analyze some external stimulus or as an indicator of an internal physiological or psychological state. That is, behavioural bioassays use behaviour to measure an animal's ability to detect environmental, physiological or neurological stimuli [21]. The behavioural sensitivity of an organism can be used as a detection device for a substance (X) in a sample [22]. Some examples of behavioural bioassays are to detect the effects of drugs, genetic mutations and chemical stimuli on the swimming behaviour of ciliates such as *Tetrahymena* and *Paramecium* [23]; to detect environmental pollutants using fish swimming behaviour [24] [25] to detect urea in the blood using a rat's maze performance [26] or to detect sex pheromones by measuring moth wing flapping behaviour [27]. Behavioural bioassays are used to measure behaviour elicited by hormones and can provide information on the concentration or potency of the hormone [28]. Behavioural bioassays can also be used to detect endocrine disrupting chemicals [29]. Within neuroscience, there are a number of types of behavioural bioassays for drug action, brain lesions and neurological disorders [30] [31] [32] [33] [34].

3. How to conduct a behavioural bioassay

Like chemical assays and bioassays, a behavioural bioassay requires a standardized procedure, an "apparatus", a method for detecting the presence of (X) and quantifying the amount of (X) in a sample, and methods for controlling confounding variables and detecting errors. As behavioural testing techniques become more refined, behavioural bioassays can become specific to particular behaviours. Designing and using a behavioural bioassay requires careful consideration. The behaviours exhibited by the species being used for the assay must be established and broken down into behaviour units in order to make the behaviour objectively measurable. This involves considerations of the level of detail necessary to answer the research question(s) being asked, and qualitatively and objectively defining the behaviour being measured to differentiate it from similar behaviours the species might exhibit.

3.1. Ethograms and partial ethograms

A behavioural bioassay requires knowledge of the behavioural repertoire of the animal to be studied. An ethogram is a comprehensive collection of descriptions representing the behaviour patterns of a particular species [35]. This is a catalogue describing the discreet behaviour patterns that form that species behavioural repertoire [36]. When the ethogram includes all of the species typical behaviours it is considered a full ethogram. For example, Eisenberg [37] provides a full ethogram for *Peromyscus*. A partial ethogram is a description of the behaviours in a specific category of behaviour such as maternal behaviour, grooming behaviour, sexual behaviour, or aggressive behaviour. Partial ethograms are used when the research questions being asked are concerned only with one category of behaviour. Brown [38] used this approach to compare the maternal behaviour of C57BL/6J and DBA/2J mice through analysis of five distinct maternal behaviours. Walker and Archer [39] provide a partial ethogram for mosquito grooming behaviour, including 12 different behaviours that can be grouped into five distinct sequences. Kalueff and Tuohimaa [40] examined six components of grooming behaviour in mice. The reproductive behaviour of pigeons was observed and catalogued by Fabricius and Jansson [41], who identified 19 important and distinct behaviours relating to reproductive behaviour. Tinbergen [42] analyzed the reproductive behaviour of the stickleback fish, from the initial defense of its territory and nest building through to the hatching of its eggs. Similarly, Mertz and Barlow [43] identified four distinct reproductive behaviours, and five motor patterns in the flagfish. Brown and McFarland [44] studied rat sexual behaviour by recording behaviours relating to mounting, ejaculating, and intromission. Some ethograms include more detailed behavioural categories than others, and are therefore a more precise measure of behaviour [36].

3.2 Qualitative description of behaviour units

Behaviour units are the behaviours that are measured in an ethogram. Eisenburg [37] described these units as stereotyped muscular coordination that can be organized into sequences and predictably elicited by either internal or external stimuli. They must be objectively defined and specific enough that different observers would record the behaviour in the same manner [45]. Behaviour units break up the continuous stream of behaviour into a series of well-defined discreet units that can be identified and quantified. Defining these behaviour units is a crucial preliminary step to conducting any behavioural experiment. The terminology and concepts used to describe behaviour have been the topic of much debate, and have evolved since the 1950's. The concept of a fixed action pattern (FAP) was one of the first terms widely adopted by behavioural researchers [46]. A FAP is generally described as a stereotyped behaviour that is not learned and is species characteristic, although there are many differing perspectives on how this term should be defined. Dewsbury [47] attempted to define FAPs unambiguously by analyzing their use in ethological research. He found many defining criteria of FAPs, but no consensus amongst his sources as to which criteria should be included. Barlow [46] also recognized difficulties associated with the notion of FAPs being used in research, namely that the concept implies a high level of stereotypy and other properties that are not realistic in practice. He proposed "modal action patterns" (MAPs) be used in place of FAPs and defined them as distinguishable spatiotemporal patterns of movement that can be named and statistically characterized, cannot be subdivided into independent units, and that are widely distributed throughout an interbreeding population. We call these modal action patterns "behavioural units".

Describing behaviour units qualitatively requires a verbal definition that is objective and allows for consistency of measurement. This can be achieved by giving unambiguous, comprehensive definitions of the behaviours being studied [36]. Units can be described in terms of structure, which describes the animal's appearance such as posture or movements; or consequence, which involves the effect of the behaviour such as escaping from a predator. Eisenberg [37] gives a list of the postures and movement patterns for adult *Peromyscus,* accompanied by further qualitative descriptions of the behaviours. Illustrations are often provided to aid in the descriptions of behaviour units, as done by Tinbergen [42] to describe stickleback fish reproductive behaviours. Walker and Archer [39] use iluustrations of each behaviour unit in their study of mosquito grooming behaviour. Benjamini et al. [47] discuss some of the issues surrounding descriptions of animal behaviour and give 10 suggestions for improving the quality of such descriptions. The publication of videos to describe behaviour units can also be done through the Journal of Visualized Experiments (JOVE) and some journals allow for video supplements to papers

3.3. Quantitative description of behaviour units

Once the stream of behaviour is divided into well-defined behaviour units, these units can be measured quantitatively. The frequency of a behaviour unit within a sample, the mean and total duration of a behaviour unit within a sample can be calculated, or the interval between displays of a particular behaviour can be measured. Brown and McFarland [44] characterized sexual behaviours of male rats through the use of these different measures, such as ejaculation frequency, bout duration, and inter-intromission-interval, resulting in the ability to more completely analyze the behaviours. Latency to onset of a particular behaviour can also be measured. This approach was used by Lee and Brown [49] to compare the latency of mice to approach pups before and after brain lesions, and in Brown and McFarland's [44] rat sexual behaviour study to measure mount latency, intromission latency, and ejaculation latency. Measuring the latency to onset of a behaviour can give the experimenter valuable insight into behavioural changes between different test conditions that otherwise would not be revealed by the data if only the occurrence of the behaviour, or frequency of the behaviour was measured. The temporal pattern of behavioural units can also be analyzed, which can lead to the discovery of patterns within a sequence. This method was used to determine that the order of song notes in chick-a-dee calls results in many distinct call types [50]. Fentress [51] used this method to assess patterns of movement in inbred mice, while Walker and Archer [39] identified five distinct grooming sequences displayed in mosquito grooming behaviour. Chatfield and Lemon [52] compared two techniques, the chi square test and an information theory approach for studying behavioural sequences in bird song.

3.3.1. Sampling and recording rules

Behaviour occurs 24 hours/day, 7 days a week. Researchers do not have time to analyze so much behaviour. Thus behaviour must be sampled for shorter periods, such as 1 hour/day for N days or in some other way. Once the time of day and duration of the behaviour sample is determined, one must decide on sampling rules and recording rules [36] Sampling rules determine which subjects will be observed and when. Using Ad libitum sampling, any behaviours exhibited by the subject that may be relevant are recorded, making it particularly useful for preliminary observations. Focal sampling records the behaviours of one individual, and is often used in studies involving groups. Weinstein [53] used this method to record ghost crab behaviour. Another method for sampling behaviour in groups is through use of scan sampling, which involves sampling the entire group for a particular behaviour at a given time. Mitlöhner et al. [54] used this method to analyze behaviours of feedlot cattle. Another method that may be chosen when observing groups is behaviour sampling, which involves recording each occurrance of a particular behaviour. In contrast to scan sampling, behaviour sampling is generally used for recording events that are less common such as copulations. Once sampling rules have been established, recording rules must also be considered.

Recording rules involve choosing whether to record every behaviour unit in a sample (continuous recording), or to use a form of time sampling (instantaneous sampling or one-zero sampling) to collect data [36]. In continuous sampling the behaviour is measured throughout the entire trial and then analyzed. Grota and Ader [55] used this method in a study measuring maternal behaviour, where time spent with the pups was considered a display of maternal behaviour. On the other hand, using time sampling methods, behaviours can be observed at regularly spaced intervals and recorded based on whether the behaviour is present or not at that moment. Bindra and Blond [56] used this method in a study sniffing, grooming, and locomotion behaviours in rats. Mitlöhner et al. [54] discuss the validity of both continuous and time sampling techniques when applied to feedlot cattle behaviour. Arrington [57] provides a review of time sampling as applied to social behaviour.

3.3.2. Lumping vs. Splitting

Behavioural units can be lumped together under a general description of behavioural categories such as grooming, sexual, aggressive, or maternal behaviour, or split apart into smaller units. For example an observer could simply record the duration of maternal behaviour, defined as mother-pup contact [55], or could split the behaviour into more detailed behaviour units such as nest building, retrieving pups, crouching and nursing, sniffing pups, and licking pups [38]. The same distinction can be made for grooming behaviour which can be split apart into more specific behavioural units such as paw licking, nose/face grooming, head washing, body, leg grooming/scratching, and tail/genitals grooming [58].

The level of detail of analysis can be increased by splitting behaviours into smaller behaviour units. For example, Bond et al. [59] looked at nest building in two strains of mice, which itself is a behaviour unit within the maternal behaviour category. Nest building was scored based on its quality, and this revealed more information about the underlying maternal behaviours of each species than simply measuring maternal behaviour as a single category of behaviour. The level of detail recorded is decided based on the research question being asked. For example, the goal of the nest building study was to resolve discrepancies in reports on nest building by B6 and

D2 mice. Therefore, having more detailed information on this particular behaviour was necessary. If the goal were to compare overall maternal behaviour, a less detailed account of behaviour would likely suffice.

3.4. Summary: the observation and description of behaviour

Accurately describing behaviour so that it may be observed, and quantitative data collected, is a process that requires careful consideration. The starting point is generally defining the research question. This helps determine whether a full ethogram or partial ethogram of the species behaviours should be used in observing a specific set of behaviours. Next behavioural units must be qualitatively defined in a manner than is objective, allowing for increased inter-observer accuracy when observing the behaviour. These units can then be measured and recorded using either continuous or time sampling methods, and quantitatively described in terms of their frequency, duration, latency to onset, or in terms of sequences that emerge. Behaviour units may be lumped together or split into smaller components during recording, depending on the level of specificity needed to answer the research question being investigated.

4. Behavioural bioassays for external and internal stimuli

Behavioural bioassays may be used to detect substances that are found in the external environment, such as toxins in water or food, and pheromones in air. They may also be used for detecting and measuring substances that are present internally such as hormones, drugs, or neurochemicals, and for studying the effects of disease processes or trauma. Finally, behavioural bioassays may be used to analyze psychological states for which there are no specific internal or external stimuli, such as fear, anxiety, preference, learning and memory.

4.1 External stimuli

A stimulus can be described as something that arouses activity or energy in a person or animal, or that evokes a specific functional reaction in an organ or tissue. External stimuli can be either simple or complex. A simple stimulus is one such as the compound (Z)-9-tricosene which is released by the male spider *Pholcus beijingensis* to increase the likelihood of a female mating with it [60]. A complex stimulus, such as the pheromone released from female Asian longhorned beetles, involves five compounds involved in attracting mates [61]. These examples consider only olfactory stimuli with complexity arising through multiple compounds acting to stimulate a response via olfactory pathways. However, stimuli may also be complex based on their ability to stimulate multiple modalities. For example, when assessing maternal behaviour in mice, the pups are a type of complex stimulus in that they offer tactile, visual, auditory, olfactory, and gustatory cues to the mother. Each of these cues may elicit maternal behaviour, or may only cause specific maternal behaviours to be displayed [55]. Similarly, when observing the responsiveness of *C. cunicularius* to *O. arachnitiformis* to flowers, Vereecken and Schiestl [62] had to determine which floral trait, odor or visual or both, was attracting the pollinator. Through careful manipulation of the variables it was established that scent was the attractant and that visual cues were not necessary to elicit the attractant response. External stimuli inducing a fear response in a mouse may also be complex, involving olfactory, visual, and auditory cues. The fear response may be elicited by one or all of the stimuli, and cannot be assumed to come from one source unless there is evidence to suggest this is the case. Thus, when designing studies that use behavioural bioassays for the presence of a particular external stimulus, it must first be established that the stimulus is in fact inducing the behaviour. Many stimuli have been shown to have the ability to elicit a particular behaviour, making them prime targets of the development of behavioural bioassays.

When an organism is exposed to toxins such as pollutants in water, internal processes can be altered which may lead to abnormal protein expression and behavioural changes. For example, expression of the luciferase gene in a rainbow trout liver tumor cell line can be used as a bioassay of toxins present in a sample [63] or a behavioural bioassay using fish swimming behaviour can be used to detect environmental pollutants, such as pesticides and metals, in water [24] [25]. Many chemical assays have been developed for detecting toxins in food such as okadaic acid in shellfish [64] and fusarium T-2 toxin in wheat [65]. Behavioural bioassays are also available to detect toxins in food such as domoic acid, a neurotoxin that causes amnesic shellfish poisoning. In this case rodent scratching behaviour can be quantified and used as an assay for the toxin [66].

Moth wing flapping behaviour can be used as a behavioural bioassay for sex pheromones present in the air [27]. The behavioural response is stereotyped in the males and can serve as a sensitive behavioural bioassay for detecting and quantifying female pheromone concentrations [67]. In this case, the proportion of males exhibiting wing flapping behaviour is positively correlated with the concentration of pheromone present. When conducting behavioural bioassays for external stimuli, it is important to consider not only what stimulus is eliciting a response but also within what range the animal can sense the stimulus. This is commonly referred to as signal detectability [68]. Behavioural bioassays have been used to measure visual ability [69] olfaction [70] and other sensory processes in mice [32].

4.2. Internal stimuli

Internal stimuli include substances such as hormones, drugs, and neurochemicals. Methods for detecting these types of stimuli may use chemical assays, bioassays, or behavioural bioassays. Behavioural bioassays can be used to detect hormone levels, drug effects, the functional effects of brain damage, and to study psychological changes in anxiety, pain, learning and memory.

It is possible to assay for hormones using both bioassays and behavioural bioassays. For example, testosterone is a hormone associated with the development of secondary sexual characteristics as well as sexual and aggressive behaviours in males [28]. A chemical assay, such as a radioimmunoassay (RIA), can be used to measure testosterone in the blood [9] [71]. Alternatively, deer antler size can be used as a bioassay for testosterone [72] or aggressive behaviour in various male species can be used as a behavioural bioassay for testosterone [73] [74] [75]. Estrogen is associated with the development of secondary sexual characteristics in females, as well as reproductive and parental behaviours [28]. The enzyme linked immunosorbent assay (ELISA) can be used to detect estrogen metabolites in urine [76]. Alternatively, cell proliferation can be used as a bioassay for estrogen [77] or walking behaviour can be used as a behavioural bioassay for estrus [78]. Prolactin is a pituitary hormone that plays a role in growth and metabolism, as well as reproductive and parental behaviours [28]. Hwang et al. [79] used a RIA as an assay for prolactin levels in primates. Bioassays for prolactin are also available, such as lymphoma cell culture proliferation [80] and the pigeon crop-sac assay [81] [82]. Prolactin is also implicated in maternal behaviour in rats [83] and paternal fanning behaviour in sticklebacks [84] indicating the potential for developing behavioural bioassays for prolactin.

Behavioural pharmacology studies the behavioural changes that are produced by drugs. These changes are largely a result of drugs modulating the excitability of synapses within neural circuits, such as the mesolimbic pathway, which is implicated in addiction [85]. Classical conditioning paradigms are often used in drug studies, where the drug acts as the unconditioned stimulus that produces internal changes in the subjects. For example, in a conditioned cue preference (CCP) test, animals are injected with a drug in an environment containing neutral cues. The animals are subsequently given a choice between the cues they experienced when the drug was administered and another set of neutral cues. The behaviour of the animal can then be used to infer the internal state that was produced by the drug; approaching the cues that were paired with the drug suggests a rewarding internal state, while avoidance suggests an aversive state [85]. This method has been used to study drugs of abuse such as amphetamine, which has been shown to induce a dose dependent CCP in rats [86]. Bardo et al. [87] conducted a meta-analysis of CCP studies that involved morphine, heroin, amphetamine and cocaine and discussed the methodological variables that can change the CCP effect size. Another common use for behavioural bioassays in pharmacology is in assessing the effect of drugs on learning and memory. These studies generally involve training an animal on a memory task, such as the radial arm maze or Morris water maze, and subsequently injecting the animal with a drug that could potentially facilitate or inhibit cellular processes involved in memory [86]. Gallagher et al. [88] trained Sprague-Dawley rats on the radial arm maze to assess the effect of opiate antagonists administered post training, and found evidence suggesting that the antagonists facilitated the rats' performance on the memory task.

Brain lesions can produce behavioural changes that are related to the function of the lesioned area. When the site of the lesion is known, such as in experimental studies, changes in the subjects behaviour can be analyzed and associated with the particular lesion site. These altered behaviours can then serve as an assay for the loci and extent of the damage present in clinical settings, or in assays for toxins, drugs, or other substances that may have neurodegenerative effects. For example, the nigrostriatal pathway, which uses dopamine as a neurotransmitter and is involved with motor control, produces motor deficits when it is lesioned, which may occur naturally as part of the Parkinson's disease process, or experimentally using the neurotoxin 6-hydroxydopamine. Schwarting and Huston [89] review the behavioural deficits such as turning, feeding, and paw usage that are produced with various lesions in dopaminergic pathways using 6-hydroxydopamine, and discuss lesion size predictions based on the results of these behavioural bioassays. Numerous behavioural bioassays can be used to analyze the effects of brain damage on motor behaviour [90] [91].

There are many psychological states for which there may be no specific external or internal stimuli that can be measured. Therse psychological states, such as attraction, aversion, fear, anxiety, pain, learning and memory, must be inferred from behavioural bioassays. It is here that the reliability and validity of the behavioural bioassay is most at risk for confounds [92]. Kalueff et al. [93] have shown how grooming behaviour can be used as a behavioural bioassay for repetitive behaviours, as occur in autism. Facial expressions have been used as behavioural bioassays for the pleasantness or unpleasantness of tastes in a taste reactivity assay [94] and for expressions of pain [95]. Behavioural tests for anxiety must be examined carefully as not all tests give the same results [96]. Finally, tests of learning and memory must be carefully analyzed in order to determine that they measure the type of learning and memory they were designed for [97].

5. Conclusion

Various types of assays are available for use in neuroscience such as chemical assays, bioassays, and behavioural bioassays. Each type is characterized by the need for carefully designed protocols involving a number of steps to be executed in order with no omissions or substitutions. In the case of behavioural bioassays this process involves deciding which behaviours to study, defining behavioural units to serve as the foundation of the assay, and deciding on when to sample and how to record the behaviours. Behaviours that are found to be driven by an internal or external stimulus can be used as an assay for that stimulus. Examples include behavioural bioassays for toxins, environmental pollutants, hormones, and brain lesions. They may be used in many industries from fisheries, to agriculture, to health sciences and are often less invasive and quicker than traditional chemical assays or bioassays making them a highly valuable tool for researchers. While behavioural bioassays for detecting sensory stimuli, either external or internal, can be verified by independent measures of the sensory stimuli, assays for psychological states cannot be so verinfied and thus must be carefully analyzed for reliability and validity.

6. References

1. Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift fur Tierpsychologie* 20, 410-433.
2. Bateson, P. and Laland, K.N. (2013). Tinbergen's four questions: an appreciation and an update. *Trends in Ecology and Evolution* 28, 712-718.
3. Nesse, R.M. (2013). Tinbergen's four questions, organized: a response to Bateson and Laland. *Trends in Ecology and Evolution* 28, 681-683.
4. Wong, A. A. and Brown. R.E. (2007). Age-related changes in visual acuity, learning and memory in C57BL/6J and DBA/2J mice. *Neurobiology of Aging* 28, 1577-1593.
5. Hinde, R.A. (1970). Animal behaviour: A synthesis of ethology and comparative psychology. Second edition. New York: McGraw-Hill.
6. Xu, S., Sun, L., & Fang, Z. (1990). Determination of gold in ore by flame atomic absorption spectrometry with flow-injection on-line sorbent extraction preconcentration. *Analytical Chimica Acta* 245, 7-11.

7. Juvonen, R., Lakomaa, T. & Soikkeli, L. (2002). Determination of gold and the platinum group elements in geological samples by ICP-MS after nickel sulphide fire assay: Difficulties encountered with different types of geological samples. *Talanta* 58, 595-603.
8. Nelson, R.J. (2005). An introduction to behavioural endocrinology. Third edition. Saunderland, Massachusetts: Sinauer. (p. 22-24).
9. Ismail, A. A., Niswender, G. D., & Midgley, A. R. (1972). Radioimmunoassay of testosterone without chromatography. *Journal of Clinical Endocrinology & Metabolism* 34, 177-184.
10. Riss, T. 2005. Selecting cell-based assays for drug discovery screening. Cell Notes, 13, 16-21.
11. Ryu, K.H., Woo, S.Y., Lee, M.Y., Jung, Y.J., Yoo, E.S., Seoh, J.Y., Kie, J.H., Shin, H.Y., Ahn, H.S. (2005). Morphological and biochemical changes induced by arsenic trioxide in neuroblastoma cell lines. *Journal of Pediatric Hematololgy/Oncology* 22, 609-621.
12. Kovalevich, J. and Langford, D. (2013). Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology. *Methods in molecular biology* 1078: 9-21.
13. Baker, M. 2015. Blame it on the antibodies. *Nature,* 521, 274-276.
14. Bradbury, A., Pluckthun, A., (2015). Standardize antibodies used in research. *Nature* 518, 27-29.
15. Begley, C.G. & Ellis, L.M. (2012). Raise standards for preclinical cancer research. *Nature* 483, 531-533.
16. Bordeaux, J., Welsh, A., Agarwal, S., Killiam, E., Baquero, M., Hanna, J., Anagnostou, V., Rimm, D. (2010). Antibody validation. *Biotechniques*,48, 197-209.
17. Pauly, D, & Hanack, K. (2015). How to avoid pitfalls in antibody use. F1000Res. 2015 Sep 7; 4: 691.
18. Plebani, M. (2010). The detection and prevention of errors in laboratory medicine. *Annals of Clinical Biochemistry* 47, 101-110.
19. Latendresse, G., & Ruiz, R.J. (2008). Bioassay research methodology: measuring CRH in pregnancy. *Biological Research for Nursing* 10, 54-62.
20. Plebani, M. (2012). Quality indicators to detect pre-analytical errors in laboratory testing. *The Clinical Biochemist Reviews* 33, 85-88.
21. Winn, P. (2001). *Dictionary of biological psychology*. New York City, New York: Taylor & Francis.
22. Braud, W. G., & Hoffman, R. B. (1973). Specificity of process, response, and stimulus in behavioral bioassays. *Journal of Comparative and Physiological Psychology* 84, 304-314.
23. Hennessey, T.M., & Lampert, T.J. (2012). Behavioral bioassays and their uses in Tetrahymena. *Methods in Cell Biology* 109, 393-410.
24. Foster, N.R., Cairns Jr, J., & Kaesler, R.L. (1969). The flagfish, Jordanella floridae, as a laboratory animal for behavioural bioassay studies. *Proceedings of the Academy of Natural Sciences of Philadelphia* 121, 129-152.
25. Scott, G.R., & Sloman, K.A. (2004). The effects of environmental pollutants on complex fish behaviour: integrating behavioural and physiological indicators of toxicity. *Aquatic Toxicology* 68, 369-392.
26. Essman, W.B. (1965). Behavioral bioassay for experimentally induced uremic endotoxemia in rats. *Perceptual and Motor Skills* 20, 115-120.
27. Jacobson, M., Adler, V.E., Kishaba, A.N., & Priesner, E. (1976). 2-Phenylethanol, a presumed sexual stimulant produced by the male cabbage looper moth, Trichoplusia ni. *Cellular and Molecular Life Sciences* 32, 964-966.
28. Wilkinson, M. and Brown, R.E. (2015). Methods for the study of behavioural neuroendocrinology. In *An introduction to neuroendocrinology*. Second edition. Cambridge, England: Cambridge University Press.
29. Clotfelter, E.D., Bell, A.M., & Levering, K.R. (2004). The role of animal behaviour in the study of endocrine-disrupting chemicals. *Animal Behaviour* 68, 665–676.
30. Buccafusco, J.J. (2001).Methods of behavior analysis in neuroscience. Boca Raton FL: CRC Press.
31. Crawley, J.N. (2007). What's wrong with my mouse?: Behavioral phenotyping of transgenic and knockout mice. Second Edition. Hoboken, NJ: Wiley.
32. Crusio, W. E., Sluyter, F., Gerlai, R.T. and Pietropaolo, S. (2013). Behavioral genetics of the mouse: Volume 1. Genetics of behavioral phenotypes. Cambridge University Press.
33. Wahlsten, D. (2011). Mouse behavioral testing: How to use mice in behavioral neuroscience. Amsterdam: Elsevier.

34. Whishaw, I.Q. and Kolb, B. (2005). The behavior of the laboratory rat. A handbook with tests. Oxford University Press.
35. Lehner, P. N. (1979). *Handbook of ethological methods*. Cambridge, England: Cambridge University Press.
36. Martin, P., & Bateson, P. (2007). *Measuring behaviour: An introductory guide*. Cambridge, England: Cambridge University Press.
37. Eisenberg, J.F. (1968). Behaviour Patterns. In J.A. King (Ed.), *Biology of Peromyscus (Rodentia*). Washington, D.C.:A.I.B.S.
38. Brown, R.E., Mathieson, W.B., Stapleton, J., & Neumann, P.E. (1999). Maternal behaviour in female C57BL/6J and DBA/2J inbred mice. *Physiology & Behaviour* 67, 599-605.
39. Walker, E.D., & Archer, W.E. (1988). Sequential organization of grooming behaviours of the mosquito, Aedes triseriatus. *Journal of Insect Behaviour* 1, 97-109.
40. Kalueff, A. V., & Tuohimaa, P. (2004). Grooming analysis algorithm for neurobehavioural stress research. *Brain Research Protocols* 13, 151-158
41. Fabricius, E., & Jansson, A. (1963). Laboratory observations on the reproductive behaviour of the pigeon (Columba livia) during the pre-incubation phase of the breeding cycle. *Animal Behaviour* 11, 534-547.
42. Tinbergen, N. (1952). The curious behaviour of the stickleback. *Scientific American, December*.
43. Mertz, J.C., & Barlow, G.W. (1966). On the reproductive behaviour of jordanella floridae (pisces: Cyprinodontidae) with special reference to a quantitative analysis of parental fanning. *Zeitschrift Für Tierpsychologie 23*, 537-554.
44. Brown, R. E., & McFarland, D. J. (1979). Interaction of hunger and sexual motivation in the male rat: A time-sharing approach. *Animal Behaviour* 27, 887-896.
45. Bart, J., Fligner, M. A., & Notz, W. I. (1998). Sampling and statistical methods for behavioural ecologists. Cambridge, England: Cambridge University Press.
46. Barlow, G.W. (1977). Modal action patterns. In T.A. Sebeok (Ed.), *How animals communicate.* Bloomington & London: Indiana University Press
47. Dewsbury, D. (1978). What is (was?) the fixed action pattern? *Animal Behaviour* 26, 310-311.
48. Benjamini, Y., Lipkind, D., Horev, G., Fonio, E., Kafkafi, N., & Golani, I. (2010). Ten ways to improve the quality of descriptions of whole-animal movement. *Neuroscience & Biobehavioural Reviews* 34, 1351-1365.
49. Lee, A. W., Brown, R.E. (2002). Medial preoptic lesions disrupt parental behaviour in both male and female California mice (Peromyscus californicus). *Behavioural Neuroscience* 116, 968-975.
50. Hailman, J.P., Ficken, M.S., & Ficken, R.W. (1985). The ‘chick-a-dee’calls of Parus atricapillus: a recombinant system of animal communication compared with written English. *Semiotica* 56, 191-224.
51. Fentress, J. C. (1992). Emergence of pattern in the development of mammalian movement sequences. *Journal of neurobiology* 23, 1529-1.556
52. Chatfield, C., & Lemon, R.E. (1970). Analysing sequences of behavioural events. *Journal of Theoretical Biology* 29, 427-445.
53. Weinstein, R.B. (1995). Locomotor behaviour of nocturnal ghost crabs on the beach: Focal animal sampling and instantaneous velocity from three-dimensional motion analysis. *Journal of Experimental Biology* 198, 989-999.
54. Mitlohner, F., Morrow-Tesch, J., Wilson, S., Dailey, J., & McGlone, J. (2001). Behavioural sampling techniques for feedlot cattle. *Journal of Animal Science* 79, 1189-1193.
55. Grota, L.J., & Ader, R. (1969). Continuous recording of maternal behaviour in Rattus norvegicus. *Animal Behaviour* 17, 722-729.
56. Bindra, D. & Blond (1958). A time sample method for measuring general activity and its components. *Canadian Journal of Psychiatry* 12, 74-76.
57. Arrington, R. E. (1943). Time sampling in studies of social behaviour: a critical review of techniques and results with research suggestions. *Psychological Bulletin* 40, 81-96.
58. Canavello, P., Cachat, J., Kalueff, A.V. (2009). Behavioural phenotyping of mouse grooming and barbering. In W.E. Crusio, F. Sluyter, & R.T. Gerlai (Eds.), *Handbook of behavioural genetics of the mouse. vol. 1: Genetics of behavioural phenotypes*. Cambridge, England: Cambridge University Press.

59. Bond, T. L., Neumann, P. E., Mathieson, W. B., & Brown, R. E. (2002). Nest building in nulligravid, primigravid and primiparous C57BL/6J and DBA/2J mice (Mus musculus). *Physiology & behavior* 75, 551-555.
60. Xiao, Y.H., Zhang, J. X., & Li, S. Q. (2010). Male-specific (Z)-9-tricosene stimulates female mating behaviour in the spider Pholcus beijingensis. *Proceedings of the Royal Society Biological Sciences* 277, 3009-3018.
61. Zhang, A., Oliver, J. E., Chauhan, K., Zhao, B., Xia, L., & Xu, Z. (2003). Evidence for contact sex recognition pheromone of the Asian longhorned beetle, Anoplophora glabripennis (Coleoptera: Cerambycidae). *Naturwissenschaften* 90, 410-413.
62. Vereecken, N. J., & Schiestl, F. P. (2009). On the roles of colour and scent in a specialized floral mimicry system. *Annals of Botany* 104, 1077-1084.
63. Richter, C. A., Tieber, V. L., Denison, M. S., & Giesy, J. P. (1997). An in vitro rainbow trout cell bioassay for aryl hydrocarbon receptor-mediated toxins. *Environmental Toxicology and Chemistry* 16, 543-550.
64. Vieytes, M., Fontal, O., Leira, F., de Sousa, J., & Botana, L. (1997). A fluorescent microplate assay for diarrheic shellfish toxins. *Analytical Biochemistry* 248, 258-264.
65. De Saeger, S., & Van Peteghem, C. (1996). Dipstick enzyme immunoassay to detect Fusarium T-2 toxin in wheat. *Applied and Environmental Microbiology* 62, 1880-1884.
66. Iverson, F., Truelove, J., Nera, E., Tryphonas, L., Campbell, J., & Lok, E. (1989). Domoic acid poisoning and mussel-associated intoxication: Preliminary investigations into the response of mice and rats to toxic mussel extract. *Food and Chemical Toxicology* 27, 377-384.
67. Shorey, H.H., Gaston, L.K., & Fukuto, T.R. (1964). Sex Pheromones of Noctuid Moths. 1. A Quantitative Bioassay for the Sex Pheromone of Trichoplusia ni (Lepidoptera: Noctuidae). *Journal of Economic Entomology* 57, 252-254.
68. Bialek, W. (1987). Physical limits to sensation and perception. *Annual Review of Biophysics and Biophysical Chemistry* 16, 455-478.
69. Wong, A.A., & Brown R.E. (2006). Visual detection, pattern discrimination and visual acuity in 14 strains of mice. *Genes, Brain and Behavior* 5, 389-403.
70. Roddick, K.M., Roberts, A.D., Schellinck, H.S., & Brown, R.E. (2016). Sex and genotype differences in odour detection in the 3xTg-AD and 5XFAD mouse models of Alzheimer's disease at 6 months of age. *Chemical Senses,* in press.
71. Lee, V.W.K., De Kretser, D.M., Hudson, B., & Wang, C. (1975). Variations in serum FSH, LH and testosterone levels in male rats from birth to sexual maturity. *Journal of Reproduction and Fertility* 42, 121-126.
72. Sempéré, A., & Lacroix, A. (1982). Temporal and seasonal relationships between LH, testosterone and antlers in fawn and adult male roe deer. *Acta Endocrinologica* 99, 295-301.
73. Book, A.S., Starzyk, K.B., & Quinsey, V.L. (2001). The relationship between testosterone and aggression: A meta-analysis. *Aggression and Violent Behaviour* 6, 579-599.
74. Gaines, M., Fugate, C., Johnson, M., Johnson, D., Hisey, J., & Quadagno, D. (1985). Manipulation of aggressive behaviour in male prairie voles (Microtus ochrogaster) implanted with testosterone in silastic tubing. *Canadian Journal of Zoology* 63, 2525-2528.
75. Wingfield, J. C., Ball, G. F., Dufty, A. M., Hegner, R. E., & Ramenofsky, M. (1987). Testosterone and aggression in birds. *American Scientist* 75, 602-608.
76. Bradlow, H.L., Sepkovic, D., Klug, T., & Osborne, M. (1998). Application of an improved ELISA assay to the analysis of urinary estrogen metabolites. *Steroids* 63, 406-413.
77. Welshons, W.V., Rottinghaus, G.E., Nonneman, D.J., Dolan-Timpe, M., & Ross, P.F. (1990). A sensitive bioassay for detection of dietary estrogens in animal feeds. *Journal of Veterinary Diagnostic Investigation* 2, 268-273.
78. López-Gatius, F., Santolaria, P., Mundet, I., & Yániz, J. (2005). Walking activity at estrus and subsequent fertility in dairy cows. *Theriogenology* 63, 1419-1429.
79. Hwang, P., Guyda, H., & Friesen, H. (1971). A radioimmunoassay for human prolactin. *Proceedings of the National Academy of Sciences* 68, 1902-1906.

80. Murphy, P.R., Friesen, H.G., Brown, R.E., & Moger, W.H. (1989). Verification of NB2 lymphoma cell bioassay for the measurement of plasma and pituitary prolactin in the Mongolian gerbil (Meriones unguiculatus). *Life sciences* 45, 303-310.
81. Simkin, B., & Goodart, D. (1960). Preliminary observations on prolactin activity in human blood. *Journal of Clinical Endocrinology & Metabolism* 20, 1095-1106.
82. Nicoll, C. S. (1967). Bio-assay of prolactin. analysis of the pigeon crop-sac response to local prolactin injection by an objective and quantitative method. *Endocrinology* 80, 641-655.
83. Loundes, D. D., & Bridges, R. S. (1986). Length of prolactin priming differentially affects maternal behaviour in female rats. *Biology of reproduction* 34, 495-501.
84. De Ruiter, A.J.H., Wendelaar Bonga, S.E., Slijkhuis, H., & Baggerman, B. (1986). The effect of prolactin on fanning behaviour in the male three-spined stickleback, Gasterosteus aculeatus. *General and Comparative Endocrinology* 64, 273-283.
85. White, N.M., & Salinas, J.A. (1998). Pharmacological approaches to the study of learning and memory. In J. L., Martinez Jr & R. P., Kesner (Eds.), *Neurobiology of learning and memory*. San Diego, California: Academic Press.
86. White, N. M., Messier, C., & Carr, G. D. (1987). Operationalizing and measuring the organizing influence of drugs on behaviour. In M.A. Bozarth (Ed.), *Methods of assessing the reinforcing properties of abused drugs* (p. 591-617). Springer New York.
87. Bardo, M. T., Rowlett, J. K., & Harris, M. J. (1995). Conditioned place preference using opiate and stimulant drugs: a meta-analysis. *Neuroscience & Biobehavioral Reviews* 19, 39-51.
88. Gallagher, M., King, R. A., & Young, N. B. (1983). Opiate antagonists improve spatial memory. *Science* 221, 975-976.
89. Schwarting, R., & Huston, J. (1996). The unilateral 6-hydroxydopamine lesion model in behavioural brain research: Analysis of functional deficits, recovery and treatments. *Progress in Neurobiology* 50, 275-331.
90. Brooks, S.P., & Dunnett, S.B. (2009). Tests to assess motor phenotype in mice: a user's guide. *Nature Reviews Neuroscience* 10, 519–529.
91. Stover, K.R., Campbell, M.A., Van Winssen, C.M., & Brown R.E. (2015). Analysis of motor function in 6-month-old male and female 3xTg-AD mice. *Behavioural Brain Research* 281, 16-23.
92. Schellinck, H.M., Cyr, D.P., & Brown, R.E. (2010). How many ways can mouse behavioral experiments go wrong? Confounding variables in mouse models of neurodegenerative diseases and how to control them. *Advances in the Study of Behavior* 41, 255-366.
93. Kalueff, A.V., Stewart, A.M., Song, C., Berridge, K.C., Graybiel, A.M., Fentress, J.C. (2015). Neurobiology of rodent self-grooming and its value for translational neuroscience. *Nature Reviews Neuroscience* 17, 45-59.
94. Berridge, K.C. (2000). Measuring hedonic impact in animals and infants: microstructure of affective taste reactivity patterns. *Neuroscience & Biobehavioral Reviews* 24, 173-198.
95. Langford, D.J., Bailey, A.L., Chanda, M.L., Clarke, S.E., Drummond, T.E., Echols, S., Glick, S., Ingrao, J., Klassen-Ross, T., Lacroix-Fralish, M.L., Matsumiya, L., Sorge, R.E., Sotocinal, S.G., Tabaka, J.M., Wong, D., van den Maagdenberg, A.M., Ferrari, M.D., Craig, K.D., Mogil, J.S. (2010). Coding facial expressions of pain in the laboratory mouse. *Nature Methods* 7, 447-449.
96. O'Leary, T.P., Gunn, R.K., & Brown, R.E. (2013). What are we measuring when we test strain differences in anxiety in mice? *Behavior Genetics* 43, 34-50.
97. O'Leary, T.P., & Brown, R.E. (2012). The effects of apparatus design and test procedure on learning and memory performance of C57BL/6J mice on the Barnes maze. *Journal of Neuroscience Methods* 203, 315-324.

Extending Frequent Subsequence Mining Techniques to Study Characteristic Patterns in Social Interactions

Nikolai W. F. Bode[1], Andrew Sutton[2], Lindsey Lacey[2], John G. Fennell[2] and Ute Leonards[2]

[1]Department of Engineering Mathematics, University of Bristol, Bristol, UK. nikolai.bode@bristol.ac.uk
[2]School of Experimental Psychology, University of Bristol, Bristol, UK. andysutton141@gmail.com, linds.lacey.2012@my.bristol.ac.uk, john.fennell@bristol.ac.uk, ute.leonards@bristol.ac.uk

When you hand someone a cup of coffee, does this always happen in exactly the same way? Probably not, but some basic patterns of behaviour are likely to occur again and again. For example, people may look at an object when handing it to another person. Our daily lives are full of such social interactions and understanding patterns in social interactions is therefore important for our general understanding of sociality in humans [1] and animals [2], but also for applications in behavioural diagnostics (e.g. autism [3]) or in the design of assistive robots [4,5].

We present a novel approach which extends ideas from data mining and bioinformatics [6,7] to study characteristic patterns in social interactions. Research in psychology and cognitive neuroscience has mainly focussed on general cognitive mechanisms relevant to social interactions [8,9]. We focus on detecting, characterising and comparing interaction patterns across contexts. In contrast to previous work that investigates the duration for which people show certain behaviours or the time intervals between different actions (e.g. [10,11]), we investigate regular patterns in the temporal sequence of actions of two interacting people. Instead of focussing on particular actions (e.g. sequential actions; turn-taking [11], gaze-following [12]) or interaction partners [5], our approach takes account of all interaction partners and it detects both sequential and synchronous actions. The original technique that our approach is based on, frequent subsequence mining, has been used to study frequent patterns in many different application areas, such as the order of items bought on a website or popular patterns in visited locations inferred from mobile phone data [13]. We argue that extending this approach to study patterns in social interaction data presents a promising avenue for future research.

We demonstrate our approach on experimental data for a simple and ubiquitous social interaction where one person hands an object (a cup) to another person (Figure 1A). We recorded a total of 440 handover interactions, 220 each in two experimental treatments (handover with an empty and a full cup). Our approach is based on representing salient features of each social interaction as a sequence of behavioural states for each individual (*behavioural state sequence*, BSS; Figure 1B). In this way, our analysis focusses on sequential or synchronous orderings of behavioural states rather than their duration or timing. Subsequently, we search for patterns that occur frequently in the 440 BSS we record (inspired by [6]), obtaining a set of *frequent sub-sequences* (FSs; Figure 1C). By selecting the most informative FSs, we obtain a set of behavioural sequences that characterise the interaction. We call these sequences *characteristic interaction sequences* (CIS) and use them in our analysis of the data. We study properties of the CISs in general, to investigate defining features of the interaction, and we compare the occurrence frequency of CISs across interaction contexts (inspired by [7]). We implement our analysis in custom-made JAVA™ (www.java.com/) software which is freely available from the lead author on request.

Ethical approval for our experiment was granted by the Ethics Committee of the University of Bristol and we obtained informed consent in writing from all participants.

Our results show that even the simple social interaction of one person handing a cup to another is characterised by a large number of distinct CISs. Many of these do not simply relate to the physical handover, but capture changes in attention during different stages of the handover. The majority of the CISs remain characteristic of the interaction when the nature of the task is manipulated by adjusting how full the cup is. However, we also identify CISs that are significantly more likely to occur in interactions with full cups. Comparing the occurrence frequency of CISs in interactions across participant pairs reveals that our handover interactions fall into three clusters: handovers specific to empty cups, handovers specific to full cups and handovers that are marked by peoples' focus and coordination of visual attention on the cup.

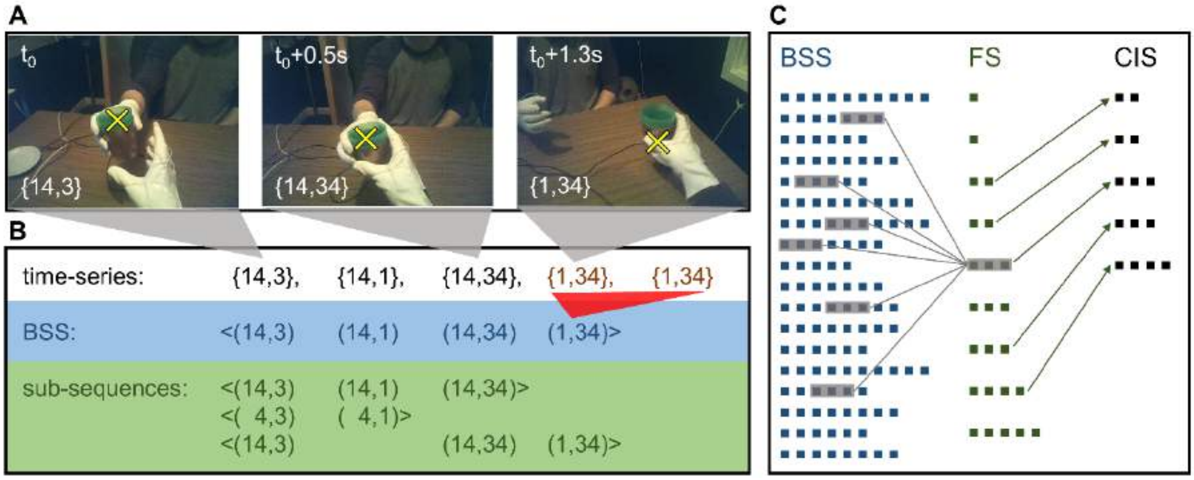


Figure 1. (A) Video snapshots from mobile eye-tracker in experiment. The gaze fixation point of the participant is marked with a yellow 'X'. Time is shown in the top left corner and behavioural states we use to represent the interaction in the bottom left corner of images (1 – look elsewhere, 2 – look at other person, 3 – look at cup, 4 – hold cup). In the interaction shown, the participant opposite holds the cup first, then both participants hold the cup and finally, only the participant whose perspective is shown holds the cup and completes the handover by placing the cup on a saucer at the other end of the table. We always write the behavioural states for the participant who initiates the handover to the left of the comma that separates the behavioural states of individuals. (B) We represent the interactions as time series of behavioural states (not shown at recording frequency of 24 Hz). As we do not focus on the duration of behaviours, but on the sequential or synchronous ordering of behavioural states, we turn these time series into behavioural state sequences (BSSs) by removing identical consecutive combinations of behavioural states (in red on first line of panel B). The symbols '<' and '>' indicate start and end of a BSS, respectively. A BSS is a sub-sequence of another BSS, if all sequence elements of the former are contained in the latter in the same order (examples in bottom three lines in panel B). We allow a gap of at most 1 sequence element in sub-sequences, to avoid recording spurious behavioural states (bottom line in panel B). (C) We search for frequent sub-sequences (FSs) that occur in a threshold number of all BSS in our data (example indicated in grey; squares indicate sequence elements). We exclude FSs of length 1 and ones that are sub-sequences of other FSs, as information in them is already contained in other FSs. In this way, we obtain a reference set of characteristic interaction sequences (CISs) which we use in the analysis of our data.

In conclusion, our results demonstrate that our approach is useful for detecting characteristic patterns in physical handover interactions. Using our approach we show that characteristic patterns exist and that, while many of them are preserved across contexts, some depend on the difficulty of the task (cup fullness) or on the individuals involved (visual attention cluster).

Our findings are directly relevant to research into human-robot interactions. One way to ensure seamless and successful interactions between machines and humans is to program machines to anticipate or predict the actions of humans based on observed action patterns (e.g. [4,5]). A characterisation of interaction patterns and knowledge of changes in patterns across contexts or even individuals is very useful for this endeavour.

The outcome of our analysis is a set of interaction sequences that occur frequently in one type of social interaction. Future work should address two questions regarding this set of sequences. First, given a new observation of a sequence of behavioural states, can the set of characteristic sequences be used to develop sufficient conditions for determining whether the newly observed interaction is a handover interaction or not? Second, our analysis framework is general and can be applied to different types of social interactions (e.g. people throwing and catching a ball). It should be investigated whether and how sets of interactions sequences that occur frequently in other types of interactions differ from the set of sequences we detected for handover interactions. Assessing the similarities or differences of sequences or sets of sequences is a well-studied problem in bioinformatics [7,14]. Adapting these techniques to our reference sets of social interaction sequences presents a highly novel way to quantitatively investigate similarities and differences between different types of social interactions.

References

1. Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters* (Vol. 7). Cambridge, UK: Cambridge University Press.
2. Witzany, G. (2014). *Biocommunication of Animals*. Heidelberg, Germany: Springer.
3. Shepherd, S.V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience* 4, 5.
4. Strabala, K.W., Lee, M.K., Dragan, A.D., Forlizzi, J.L., Srinivasa, S., Cakmak, M., Micelli, V. (2013). Towards seamless human-robot handovers. *Journal of Human-Robot Interaction* 2, 112-132.
5. Zheng, M., Moon, A., Croft, E.A., Meng, M.Q.H. (2015). Impacts of Robot Head Gaze on Robot-to-Human Handovers. *International Journal of Social Robotics* 7, 783-798.
6. Agrawal, R., Srikant, R. (1995). Mining sequential patterns. *Proceedings of Data Engineering 1995, 11th International IEEE Conference on Data Engineering* (Taipei, 6-10 March 1995), 3-14.
7. Hirschhorn, J.N., Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6, 95-108.
8. Carey, D.P., Della Sala, S., Ietswaart, M. (2002). Neuropsychological perspectives on eye-hand coordination in visually-guided reaching. *Progress in Brain Research* 140, 311-327.
9. Ferrari, P.F., Bonini, L., Fogassi, L. (2009). From monkey mirror neurons to primate behaviours: possible 'direct' and 'indirect' pathways. *Philosophical Transactions of the Royal Society B* 364, 2311-2323.
10. Yu, C., Smith, L.B. (2013). Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE* 8, e79659.
11. Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J.P., Yoon, K-E., Levinson, S. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106, 10587-10592.
12. Adams R.B., Ambady, N., Nakayama, K., Shimojo, S. (2011). *The science of social vision* (Vol. 7). Oxford, UK: Oxford University Press.
13. Gupta, M., Han, J. (2011). Applications of pattern discovery using sequential data mining. *In Pattern Discovery Using Sequence Data Mining: Applications and Studies (Kumar, P., Krishna, P.R. and Raju, S.B. eds.)*, Hershey, USA: IGI Global, pp. 1-23.
14. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24, 1586-1591.

New conditioning chambers for free flying and lab experiments with honey bees

M. Sokolowski

GRAP, Université de Picardie Jules Verne, Armiens, France. michel.sokolowski@u-picardie.fr

Skinner boxes are well used tools for behavioral and neuroscience studies. Standardized chambers are available for rodents, birds and primates. Unfortunately, despite several successfully tested prototypes (see for example, Sigurdson, 1981 or Sokolowski and Abramson, 2010), no such boxes are available for insects or honey bees. However, because a Skinner box can simultaneously measure behavior and control the environment (especially food delivery conditions) such standardized tools could be particularly useful for foraging and ecotoxicological research.

We present here a new autonomous microcontrolled device that can be used in bee operant conditioning chambers. Three versions have been tested. The first version includes one feeder and is designed to work in the lab with a group of honey bees observed continuously during several weeks with no human intervention. This version is an improvement of the classic bee feeding box using an Eppendorf tube as a feeding device. The second version, similar to the first one, includes two feeders for choice and discrimination studies. The third version has been designed to work with individual free flying nectar foraging bees. We demonstrated the usefulness of the new tool with several simple conditioning protocols and show how new results can be easily obtained in feeding, choice, or foraging experiments.

References

1. Sigurdson, J. E. (1981). Measurement of consummatory behavior in honey bees. *Behavior Research Methods & Instrumentation,* **13**, 308-310.
2. Sokolowski, M.B.C., Abramson, C.I.A. (2010). From foraging to operant conditioning: A new computer-controlled Skinner box to study free-flying nectar gathering behavior in bees. *Journal of Neuroscience Methods,* **188**, 235-242.

New Tools for Discovering and Analyzing Recurrent Real-Time Behavior Patterns: THEME 6 with BehaviorCoder and the new Theme(Smart)Watch 1

M.S. Magnusson

Human Behavior Laboratory, University of Iceland, Reykjavik. msm@hi.is

This paper concerns new tools for the measurement of the behavior and interactions of individuals as diverse as neurons and humans, through the detection and analysis of particular types of recurrent time-structured patterns, mainly T-patterns and T-packets, that together make up the "T-system". This type of analysis of the temporal structure of behavior, views the organization of real-time behavioral streams partly as T-patterns, that is, as repeated structured hierarchical self-similar clusters characterized by statistically significant translation symmetry between the occurrences of each pattern. These repeated structures can be considered a particular type of repeated statistical (pseudo) fractal objects. The T-packet, as an extension of the T-pattern, is a T-pattern that, in a particular statistical manner, attracts or repulses particular other behaviors. A comprehensive review of the many different applications of the dedicated T-Pattern Analysis (TPA) software THEMETM to the analysis of human, animal and neuronal behavior and interactions, has recently been published by J. Neuroscience Methods [1] and a different application is described in each chapter of a new edited book [2] a kind of sequel to an earlier TPA/Theme application book [3] also mostly by the Methodology for the Analysis of Social Interaction (MASI) international university research network (see hbl.hi.is).

After years of use as a beta version, a new and much more powerful THEME version 6, up to 50 times faster and accepting much bigger data, is being launched and should be available on the PatternVision website about the time of Measuring Behavior 2016, a full commercial version and a free educational version. (Marketing is now exlusively in the hands of PatternVision Ltd, so for availability and ordering see www.patternvision.com). Some of its new features will be described here. Statistical validation of its findings is now based on two different Monte Carlo methods that are also available for single patterns. Regarding TPA with THEME such validations are still particularly important as no other available software searches for this type of patterns making comparisons of results impossible. The earliest example of the typically strong both statistical and external validation of T-pattern detection with Theme concerned children's dyadic interactions, but strong validations have recently also been found in a very different study of interactions within populations of neurons in the rat olfactory [4]. While even the statistical significance of results obtained from such neuronal data using other, somewhat similar, but simpler models and detection algorithms, remain in doubt, TPA with Theme allowed abundant detection of highly significant complex multi-neuron T-patterns. Strong association with breathing provided the initial external validation, but thanks to the unexpected gradual deterioration (trials finally had to be stopped) of one of the subjects, further **external** validation was found. That is, while the spiking neurons and their spiking rate remained practically unchanged, the patterns gradually became simpler and fewer and the Monte Carlo results went gradually from extremely significant to barely significant.

All input data files now follow a simple two column tab delimited data format (time event) with one sample (observation period) per file, and Theme 6 includes tools for checking, correcting and recoding, while samples can be easily regrouped according to the experimental design thus facilitating the discovery of experimental effects.

Theme 6 allows various new kinds of analysis of the pattern composition in sets of detected T-patterns. For example, analysis and selection of detected patterns according to occurrence of particular behaviors in particular positions within patterns, the relevance of which has been shown in a recent study of deception in human interactions allowing the identification of cheaters [5]. Theme thus continues to detect effects of independent (experimental) variables easily missed by standard statistical analysis methods.

A new feature now allows prediction of a selected target behavior (a coded event-type or a detected T-pattern) on the basis of the full set of detected T-patterns in the data.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Theme 6 also allows the detection of the so called T-associates and T-packets with various new graphical possibilities for their visualization and analysis.

Discovering patterns occurring significantly more in an experimental group is now possible and easy in Theme 6. Similarly, observations can be split based on particular events or event pairs and other kinds of time intervals can be defined in various ways within the data (observation) and T-patterns discovered that have a statistically significant tendency to occur inside (or outside) these intervals.

Theme 6 can generate many new kinds of (standard tab delimited) statistical tables regarding the detected T-patterns for further analysis with statistical programs such as SPSS, CSS, SAS, R, etc.

Theme 6 and Big Data vs. "Tiny data". The largest single datasets analyzed as yet with Theme 6 consist of about 100.000 events, but hundreds of datasets each of up to 40.000 events have also been searched for T-patterns (in principle, no practical upper limit exist for their number) in one Theme run. While such data are quite big they may still be considered small compared to the, so called, Big Data. On the other hand, a methodological TPA characteristic essentially due to its particular and intensive use of real-time information, should be underlined, that is, TPA applicability to Tiny Data, even down to just a few events and typically too small for useful standard statistical analysis, but possibly providing extremely significant TPA findings.

Economical aspect for experiments. Due to the intensive use of real-time information, TPA with Theme may be economical by finding experimental effects with relatively small data and thus possibly also fewer subjects.

It should be underlined that TPA with Theme has been applied to data with temporal resolutions (smallest time unit) ranging from a few millionth of a second to a day and is clearly **not** limited to data obtained through video recordings. But, as its earliest versions, Theme now comes with a simple (digital) video coder, called BehaviorCoder, especially designed for use with Theme, and can be launched from within Theme or separately. Using the same project folder as Theme, it automatically uses the same category (vvt or variable/value) table and produces data files that can be immediately imported into Theme for analysis.

ThemeWatch 1 is a brand new SmartWatch app(lication) designed for TPA with Theme. The ThemeWatch allows the user to easily log time-stamped events in his/her behavior (life) and, when she/he wants and then **Send** the log (a Theme data file) to an email address of his/her choice for TPA with Theme. This very first version of the ThemeWatch will be presented.

References

1. Casarrubea M, Jonsson GK, Faulisi F, Sorbera F, Di Giovanni G, Benigno A, Crescimanno G, Magnusson MS. (2015). T-pattern analysis for the study of temporal structure of animal and human behavior: A comprehensive review. *Journal of Neuroscience Methods*, **239**, 34-46.
2. Magnusson, MS, Burgoon, JK, Casarrubea, M. (Eds.) (2016) Discovering Hidden Temporal Patterns in Behavior and Interaction: T-Pattern Detection and Analysis with THEME™. Neuromethods Series. Springer; 1st ed. 2016 edition (January 12, 2016).
3. Anolli L., Duncan S Jr., Magnusson MS and Riva G (eds.) (2005). The Hidden Structure of Interaction: From Neurons to Culture Patterns. IOS Press, Amsterdam.
4. Nicol AU, Segonds-Pichon A, Magnusson MS (2015). Complex spike patterns in olfactory bulb neuronal networks. *Journal of Neuroscience Methods,* **239**, 11-17.
5. Burgoon, JK. (et al.) (2016) Interactive Deception in Group Decision-Making: New Insights from Communication Pattern Analysis. In Magnusson, MS, Burgoon, JK, Casarrubea, M (Eds.) (2016) Discovering Hidden Temporal Patterns in Behavior and Interaction: T-Pattern Detection and Analysis with THEME™. Neuromethods Series. Springer; 1st ed. 2016 edition (January 12, 2016).

T-pattern Detection and Polar Coordinate Analysis of Motor Skills Used by Lionel Messi in Goal Scoring in Soccer

M. Castañer[1], D. Barreira[2], O. Camerino[1], M.T. Anguera[3], G.K. Jonsson[4], R. Hileno[1], A. Canton[1]

[1] INEFC, University of Lleida, Spain; mcastaner@inefc.es
[2]Faculty of Sport, University of Porto, Portugal
[3]Research Institute IR3C, University of Barcelona, Spain
[4]Human Behaviour Laboratory, University of Iceland, Reykjavik, Iceland

Abstract

Research in soccer has traditionally given more weight to players' technical and tactical skills, but few have analyzed the motor skills that underpin specific motor actions. The objective of this study was to investigate how one of the world's top soccer players, Lionel Messi, uses his motor skills in attacking actions that result in a goal. We used an observational methodology design and built a specific, easy-to-use observation instrument (OSMOS-goal scoring) adapted from an existing motor skills observation instrument (OSMOS). The instrument comprised 4 criteria (body part, foot contact zone, body orientation, and turn direction) and 20 exclusive, mutually exhaustive categories. Associations between these categories were investigated by T-pattern detection and polar coordinate analysis. The findings show how Messi uses his left foot much more frequently than his right foot, which is to be expected considering that he is left-footed. The T-pattern analysis showed that he angles his body towards the left side line while controlling the ball with the inside of his left foot at the beginning of all sequences leading up to a goal and then faces the goal line while controlling the ball with his left instep. These results were complemented by the polar coordinate analysis findings, which showed that Messi's angling of his body towards the goal line activates use of his right foot (mainly the outer part) while facing right, allowing him to move towards the right flank to create space to the left to take a shot with his predominant left foot.

Keywords: Motor behavior; elite players; pattering detection; vectored actions: polar coordinate analysis.

Introduction

Soccer research has traditionally focused on technical and tactical aspects of team play [1], but few studies have analyzed individual actions, such as goal scoring, and fewer still have analyzed the motor skills supporting these actions. Styles of play that incorporate and encourage individual actions and skills can improve overall game strategies and outcomes [2]. Use of motor skills in elite soccer has largely been studied from a subjective perspective [3], but mastery of these skills [4-6] is directly linked to motor versatility [5] and consequently to the ability to execute complex intentional actions [7].

Motor versatility requires the integration of multiple skills [5,8]; it is a particular important quality in attackers such as strikers and wingers, and is closely linked to perceptual anticipation [9]. These multiple skills are essential to the execution of soccer moves, such as ball control, dribbling, feints, and shots. Motor skills involve axial movements in the form of turns and pivots, spatial orientation of the player's body in relation to the side lines and goal line, and use of one limb or another (laterality). These movements not only underpin all soccer moves, but also contribute to the uniqueness of each player. In addition, most of these movements are interlinked, as laterality [5,10], for example, refers not only to left-right preference but also to how a player orientates his body spatially [5,11,12,]. Lionel Messi is a good example in relation to these last two aspects, as he is a left-footed player who has achieved some of his best results playing on the right wing.

Messi is considered to be one of the world's top soccer players and is typically described as a unique, skilled, and versatile player. These attributes, however, have not been analyzed from an objective, scientific perspective.

Such an analysis is challenging, as soccer is a complex game that requires a wide repertoire of individual skills used to the benefit of the team and is characterized by constant interactions between technical, tactical, psychological, and physical factors. Two particularly fitting techniques for analyzing such complexity are T-pattern (temporal pattern) detection [13,14] and polar coordinate analysis [15].T-pattern detection has been successfully used in numerous studies to reveal hidden patterns underlying different soccer actions [16-19]. Polar coordinate analysis, in turn, is a powerful data reduction technique that is being increasingly used in studies of team sports [20,21]. The technique provides a vectorial representation of the complex network of interrelations between carefully chosen, exhaustive and mutually exclusive categories included in an ad hoc observation instrument.

The overall objective of this study was to perform an objective analysis of Lionel Messi's use of motor skills prior to scoring a goal using two complementary methods: T-pattern detection and polar coordinate analysis. The methodological aim was to identify hidden temporal patterns underlying these skills and to provide a vectorial representation of the associations detected between behaviors using polar coordinate analysis, whose powerful data reduction feature facilitates the interpretation of data.

Methods

Observational methodology [22] offers eight types of observational designs [23,24] that have been widely applied in the analysis of team sports [17,18,25,26] and use of motor skills in physical activity and sport [4,27]. We chose the Idiographic/Follow-up/Multidimensional (I/S/M) design because our study focused on a single player (idiographic), analyzed specific moves recorded from the beginning to end of different sequences (follow-up), and addressed multiple criteria and responses included in an ad hoc observation instrument (multidimensional). The methods applied were T-pattern detection and polar coordinate analysis.

Participants

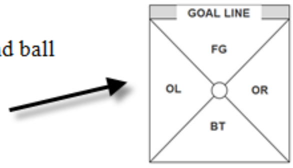
We analyzed the execution of 103 goals scored by Messi in top competitions over a decade (2004-2014). Our study can thus be considered case-oriented [28-30]. The goals were analyzed using public television footage, in compliance with the ethical principles of the Declaration of Helsinki and with approval from the ethics committee at the university overseeing the study. The goal inclusion criteria were a) clear observability of the sequence leading up to the goal [23] and b) availability of at least two recordings of each sequence from a different angle. Goals scored directly following reception of the ball were not included.

Instruments

Observational instrument. A purpose-designed observation instrument (OSMOS-goal scoring) adapted from the OSMOS motor skills observation instrument [4] was used. The instrument comprised four criteria: body part (part of the body that the players uses to make contact with the ball), foot contact zone (part of the foot used to touch the ball), body orientation (angling of the body with respect to the side line or goal line), and turn direction (right vs left). Each of these criteria was expanded to build an exhaustive, mutually exclusive observation system that included 20 categories (see Table I). A previous analysis of three additional criteria related to external aspects that might intervene in goal scoring opportunities—pitch zone, pitch side, and number of opponents the attacker needs to pass—was conducted, but no T-patterns or significant associations in the polar coordinate analysis were observed.

Table I. OSMOS-goal scoring. Observation System for Motor Skills Used in Goal Scoring

Criterion	Category	Code	Description
Body part	Left foot	LF	Player touches the ball with left foot
	Right foot	RF	Player touches the ball with right foot
	Left leg	LL	Player touches the ball with left leg (not including foot)
	Right leg	RL	Player touches the ball with right leg (not including foot)
	Chest	CH	Player touches the ball with chest
	Back	BA	Player touches the ball with back
	Head	HD	Player touches the ball with head
Foot contact zone	Tip	TI	Player touches the ball with tip of foot
	Outside	OU	Player touches the ball with outside of foot
	Inside	ID	Player touches the ball with inside of foot
	Heel	HL	Player touches the ball with heel
	Sole	SO	Player touches the ball with sole
	Instep	IT	Player touches the ball with instep
	Non-observable	NON	No clear contact zone between player and ball
Body orientation	Facing goal	FG	Player's chest facing rival goal line
	Facing right	OR	Player's chest facing right side line
	Back to goal	BT	Player's back facing rival goal line
	Facing left	OL	Player's chest facing left side line
Turn direction	Right turn	RT	Player makes a full or half turn to the right (vertical axis)
	Left turn	LT	Player makes a full or half turn to the left (vertical axis)



Recording instrument. Goals sequences were coded using LINCE (v.1.2.1) [31]. This software program was also used for the data quality check.

Data analysis software. Three programs were used: a) Theme software package [14] for T-pattern detection; b) GSEQ v5.1.15 [32] for a preliminary analysis of the data [33]; and c) HOISAN v1.3.6.1 [34] for the polar coordinate analysis.

Procedure

Goals sequences were analyzed from the moment Messi received the last pass up to the moment he scored a goal. Two expert observers, a national soccer coach and a motor skills expert, recorded the sequences following appropriate training in the use of OSMOS-goal scoring. Interobserver reliability was calculated in LINCE before coding the full data set using a preliminary dataset of 35 goal sequences not included in the final sample. The resulting kappa statistic was 0.92, which guarantees the interpretative rigor of the coding process.

Data analysis

T-pattern detection. T-pattern detection is one of the most widely used techniques in observational methodology [23]. The following criteria were applied to guarantee that any T-patterns detected were not due to random events: a) presence of a given T-pattern in at least 25% of all sequences, b) significance level of 0.005, and c) redundancy reduction setting of 90% for occurrences of similar T-patterns. As Magnusson states [13,19], the idea of T-pattern analysis is to detect repeated behavioral patterns that are invisible to unaided observers. The temporal structure of complex behavioral sequences is composed of simpler or directly distinguishable event-types. Each T-data set subject to analysis consists of series of behaviors coded as occurrence times (beginning and end points) within specified observation periods (time point series).

Within a given observation period, two actions, A and B, that are repeated, either in this same order or simultaneously, form a minimal T-pattern (AB) if they are found more often than would be expected by chance, and if, assuming the null hypothesis of independent distributions for A and B, they are separated by approximately the same distance (time). Instances of *A* and *B* separated by this approximate distance constitute an (*AB*) T-pattern and their occurrence times are added to the original data. More complex T-patterns consisting of simpler, already-detected patterns, are subsequently added through a bottom-up detection procedure. Pairs or series of patterns can thus be detected, for example (((*AB*)*C*)(*DE*)) (see our example in Figure 1). The Theme software package features algorithms for dealing with potential combinatorial explosions due to redundant and partial detection of patterns using an evolution algorithm (completeness competition), which compares all patterns and retains only the most complete ones. As any basic time unit can be used, T-patterns are, in principle, scale-independent, although only a limited range of basic unit sizes is relevant in any study.

Polar coordinate analysis. Polar coordinate analysis involves establishing significant associations between a given behavior (the behavior of interest) and conditional behaviors (the other behaviors analyzed). Before these associations can be investigated by polar coordinate analysis, however, it is first necessary to apply data reduction techniques to identify behaviors with statistically significant associations. This preliminary reduction was performed in our case by lag sequential analysis with multievent data and lags ranging from -5 to +5 in GSEQ v5.1.15 [23,32]. The results are summarized in Table II. Significant values ($p<.05$) were identified following application of the binomial test and calculation of adjusted residuals using the procedure described by Allison and Liker (1982) [35]. Both positive (>1.96) and negative (<-1.96) values were considered to be significant.

Table II. Given and conditional behaviors showing significant associations in the preliminary analysis.

Given Behavior	Conditional Behavior
LF (Body part: Left foot)	LF (Body part: Left foot)
RF (Body part: Right foot)	RF (Body part: Right foot)
RL (Body part: Right leg)	RL (Body part: Right leg)
CH (Body part: Chest)	CH (Body part: Chest)
HD (Body part: Head)	HD (Body part: Head)
OU (Foot contact zone: Outside)	OU (Foot contact zone: Outside)
ID (Foot contact zone: Inside)	ID (Foot contact zone: Inside)
NON (Foot contact zone: Non-observable)	NON (Foot contact zone: Non-observable)
FG (Body orientation: Facing goal)	FG (Body orientation: Facing goal)
OR (Body orientation: Facing right)	OR (Body orientation: Facing right)
BT (Body orientation: Back to goal)	BT (Body orientation: Back to goal)
OL (Body orientation: Facing left)	OL (Body orientation: Facing left)
RT (Turn direction: Right)	RT (Turn direction: Right)

Polar coordinate analysis involves the application of a complex procedure that uses adjusted residuals obtained from the preliminary lag sequential analysis [36] to provide a vector map of interrelated behaviors. The same number of prospective and retrospective lags are analyzed in each case. Prospective lags show which conditional behaviors precede the given behavior, while retrospective lags show which behaviors follow it [37].

Polar coordinate analysis merges the prospective and retrospective approaches to achieve a powerful reduction of data through the calculation of the *Zsum* statistic $\left(\frac{\Sigma z}{\sqrt{n}}\right)$ described by Cochran (1954) [38] and later developed by Sackett (1980). In both the prospective approach ($Z_{sum}P$) and the retrospective approach ($Z_{sum}R$), calculations are based on the frequency of the given behavior, *n*, and a series of mutually independent *z* values for each lag. Each of these values is obtained by applying the binomial test to compute conditional probabilities (based on the number of codes recorded for each goal sequence) and unconditional probabilities (due to random effects). The length of each vector is obtained from $\sqrt{(Z_{sum}P)^2 + (Z_{sum}R)}^2$, while its angle is calculated by dividing the retrospective *Zsum* arcsine by the radius (φ = arcsine of Y/radius). Prospective and retrospective Zsum values (lags 1 to 5 and lags -1 to -5, respectively) can carry a positive or negative sign; these signs determine in which quadrant the resulting vectors (behaviors) are placed. There are four possible quadrants and these indicate the relationship between the behaviors (inhibitory vs excitatory), as shown below:

Quadrant I (++). The given and conditional behaviors are mutually excitatory.
Quadrant II (- +). The given behavior is inhibitory and the conditional behavior is excitatory.
Quadrant III (- -). The given and conditional behaviors are mutually inhibitory.
Quadrant IV (+ -). The given behavior is excitatory and the conditional behavior is inhibitory.

Depending on the quadrant in which the conditional behavior is placed, the angle of the vector is transformed as follows: quadrant I ($0 < \varphi < 90$) = φ; quadrant II ($90 < \varphi < 180$) = $180 - \varphi$; quadrant III ($180 < \varphi < 270$) = $180 + \varphi$; quadrant IV ($270º < \varphi < 360º$) = $360º - \varphi$.

HOISAN v1.3.6.1 was used to calculate the prospective and retrospective adjusted residuals and the length and angle of the vectors and to produce a graphical representation of the results.

Results

We first explored the frequency of events and event sequences (Figure 1). The box in Figure 1 shows the first 25 event-types with more than 2 occurrences.

The most frequent event-types were facing goal (FG) (n=81) and facing left (OL) (n=78), use of inside of left foot while facing the left side line (LF,ID,OL) (n=58), and use of outside of left foot (LF,OU,OL) and instep of left foot (LF,IT,OL) while facing the left side line, with respective occurrences of 42 and 38. Messi used his left foot more frequently than his right foot when facing the goal, with a slight preference for use of his instep (LF,IT,FG) (n=37) over the inside (LF,ID,FG) (n=34) or outside of his foot (LF,OU,FG) (n=30).

Messi also used his right foot while facing different directions, but to a much lesser extent than his left foot. For example, he used his right foot just 11 times while facing the left side line (RF,OL) and three times while facing the right side line (RF,OR). When his body was angled towards the goal, he used the instep of his right foot (RF,IT,FG) 12 times and the inside of this foot (RF,ID,FG) 14 times.

Analysis of event-types with fewer occurences in the chart shows Messi's preference for his left foot and also that he uses other parts of his body on isolated occasions.

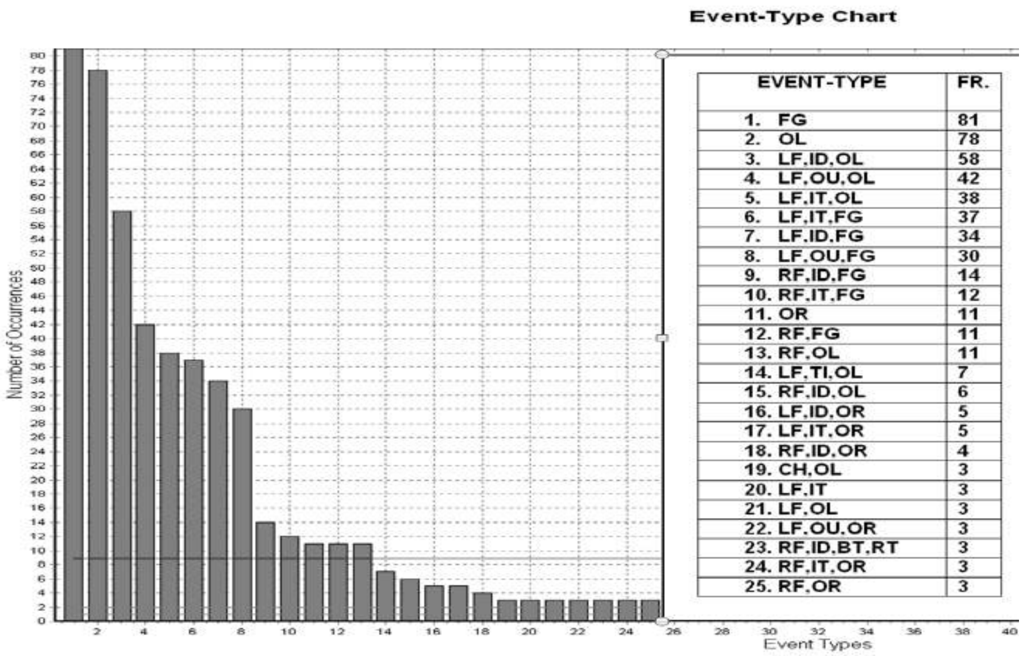


Figure 1. Event-type frequency chart.

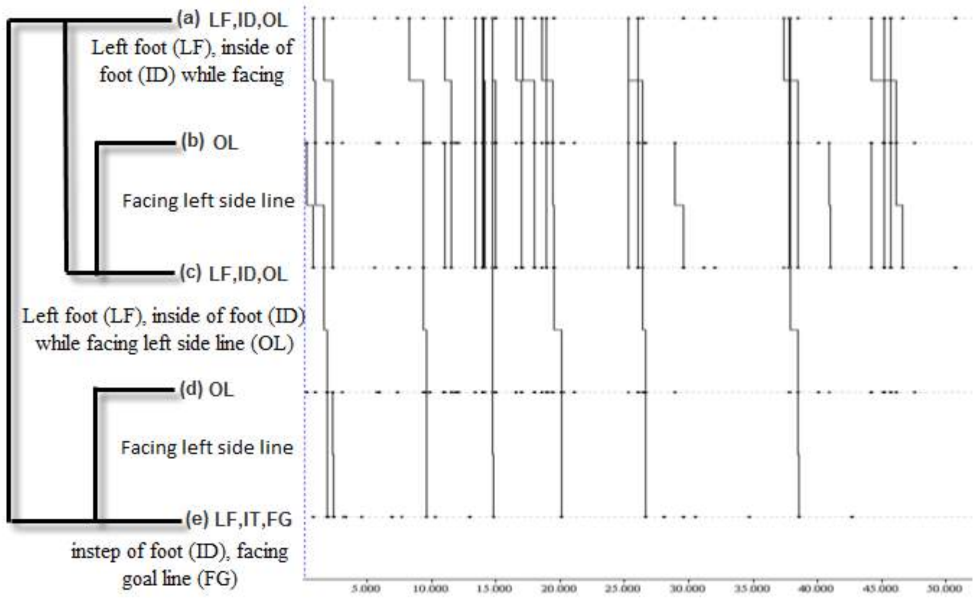


Figure 2. The T-pattern in this figure shows that (a) after Messi touches the ball with his left foot (LF) and more specifically with the inside of this foot (ID) while facing the left side line (OL), he (b) maintains this body angle (OL), (b) touches the ball again with the inside of his left foot (LF) (ID) while still facing left, (c) continues to retain this angle (OL), and (d) touches the ball again with his left foot (LF) but (e) this time with the instep (IT) and with his body facing the goal line (FG).

Category	Q	ZsumP	ZsumR	Radius	Angle
Bodypart_LF	III	-2.36	-1.6	2.85 (*)	214.27
Bodypart_RF	I	0.59	1.94	2.03 (*)	73.08
Footczone_OU	III	-1.15	-1.15	1.62	225
Footczone_ID	III	-2.33	-0.12	2.33 (*)	182.88
Bodyorient_FG	III	-3.4	-2.05	3.97 (*)	211.04
Bodyorient_OR	III	-2.91	-0.93	3.05 (*)	197.73
Bodyorient_OL	I	4.44	1.97	4.86 (*)	23.92

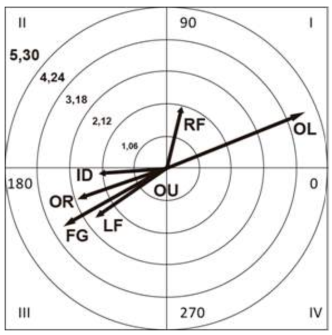


Figure 3. Map 1 shows mutual activation between Messi's use of the outside of his foot (OU) and his right foot (RF). In addition, this use of the outside of his foot (OU) activates his facing towards the left side line (OL) and, logically, inhibits use of his leg foot (LF), use of the outside of his foot (ID) and angling of his body towards the right side line (OR) and the goal (FG).

Category	Q	ZsumP	ZsumR	Radius	Angle
Bodypart_RF	III	-2.12	-1.45	2.57 (*)	214.39
Footczone_OU	I	1.97	4.44	4.86 (*)	66.08
Bodyorient_FG	III	-9.59	-9.67	13.61 (*)	225.24
Bodyorient_OR	III	-3.91	-7.38	8.35 (*)	242.07
Bodyorient_BT	II	-1.61	2.63	3.08 (*)	121.51
Bodyorient_OL	I	12.3	12.3	17.39 (*)	45
Turndirec_RT	II	-0.01	1.99	1.99 (*)	90.3

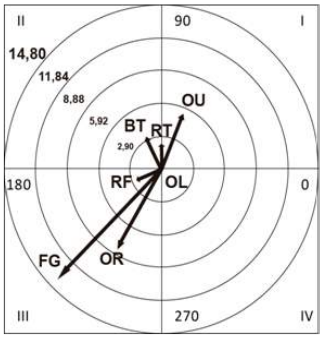


Figure 4. Map 2 shows mutual activation between Messi's angling of his body towards the left side line (OL) and use of the outside of his foot (OU); this facing left is in turn inhibited by use of his right foot (RF), facing right (OR) and facing the goal (FG).

Category	Q	ZsumP	ZsumR	Radius	Angle
Bodypart_RF	II	-0,09	2.04	2.04 (*)	92.44
Footczone_OU	III	-2.05	-3.4	3.97 (*)	238.96
Bodyorient_FG	I	8.57	8.57	12.12 (*)	45
Bodyorient_OR	I	1.02	3.85	3.98 (*)	75.21
Bodyorient_BT	IV	0.88	-1.77	1.97 (*)	296.4
Bodyorient_OL	III	.-9.67	-9.59	13.61 (*)	224.76

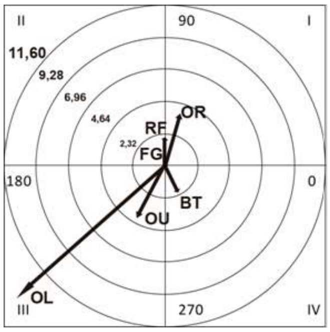


Figure 5. Map 3 shows how Messi's facing the goal line (FG) activates his facing right (OR) and inhibits his facing left (OL) and use of the outside of his foot (OU).

Discussion

Soccer actions such as ball control, dribbling, and shots at goal have been widely studied in the literature, but little attention has been given to the motor skills that support these actions. To add to the limited knowledge in this area, we systematically studied 103 goals scored by Lionel Messi and analyzed which parts of his body and feet he used to touch the ball, how he angled his body during these actions, and how he turned. Because there is evidence that overall attacking patterns of play in soccer are influenced by zone of play and interaction with defenders [i.g.16], we performed a previous analysis of numerous external factors, unrelated to motor skills, to investigate their possible influence on goal scoring by Lionel Messi. We observed no T-patterns or significant associations in the polar coordinate analysis for the three external factors included in the observation instrument (pitch zone, pitch side, and number of defenders passed). The fact that we did find significant associations for the four criteria related to motor skills (body part, foot contact zone, body orientation, and turn direction) indicates that the OSMOS-goal scoring observation instrument is particularly suited to analyzing such skills in attackers and can be incorporated as a complementary tool [22,23] in similar studies.

The merging of T-pattern detection and polar coordinate analysis proved to be a reliable strategy for 1) unveiling associations between different motor skills used in goal scoring that are undetectable by pure observation alone and 2) identifying inhibitory and excitatory relationships between the related motor behaviors following the application of a powerful data reduction technique. Examples of these results are given in Figures 2-5. The integrated vision provided by this merging of techniques showing behavioral frequencies, patterns, and associations permits an objective assessment of Messi's "uniqueness" and shows the potential of this methodology for similar studies.

Messi used his left foot much more frequently than his right foot, which is to be expected considering he is left-footed [38]. The T-pattern example given in Figure 2 shows that goals scored by Messi are typically preceded by a sequence in which he uses the inside of his left foot to control the ball while angling his body towards the left side line as he moves up the pitch and then touches the ball with his left instep while facing the goal line. Accordingly, and as shown by Map 1 generated by the polar coordinate analysis, Messi's use of the outside of his foot and his right foot are mutually activated, which is consistent with the T-pattern showing a more frequent association between use of his left foot and use of the inside and instep of this foot. Map 1 also shows how facing the left side line is activated by his use of the outer part of his foot. Consequently, the following behaviours are all mutually inhibited: use of left foot, use of the outside of foot, facing the right side line, and facing the goal.

Map 2, by contrast, shows how Messi's angling of his body towards the left side line and use of the outside of his foot to control the ball are mutually activated. His facing left is at the same time inhibited by use of his right foot and facing the right side line and the goal. Finally, map 3 shows that when Messi angles his body towards the left side line, facing the right side line and use of the outside of his foot are both inhibited. These findings show how Messi controls the ball with the outside of his right foot to move towards the right flank in order to find the space he needs to take a shot with his stronger left foot.

From a methodological perspective, we have shown the enormous potential offered by observational methodology in terms of studying goal scoring and attacking actions in soccer. We have described the use of a highly specific, purposed-designed observation instrument that can be extended or adapted for use in similar studies [40] and have also shown how the merging of T-pattern detection and polar coordinate analysis is an effective technique for systematic observational studies seeking to analyze the use of motor skills in soccer and to investigate the qualities that make certain players unique. In our analysis of 103 goals scored by one of the world's greatest ever soccer players, we detected interesting associations between different motor skills used by Lionel Messi and showed how he uses and angles his body to control the ball as he moves up the pitch in search of a goal. Both techniques are particularly suited to detecting and mapping the broad repertoire of motor skills required to become a successful player. Considering that motor skills are the "motor" behind all actions on the soccer field, we believe that more studies of this type are needed.

References

1. McGarry, T., Anderson, D.I., Wallace, S.A., Hughes, M.D., & Franks, I.M. (2002). Sport competition as a dynamical self-organizing system. *Journal of Sports Sciences*, 20(10), 771–781.
2. Carling, C., Bloomfield, J., Nelsen, L., & Reilly, T. (2008). The role of motion analysis in elite soccer: Contemporary performance measurement techniques and work rate data. *Sports Medicine*, 38(10), 839–862.
3. Duch, J., Waitzman, J.S., & Amaral, L.A.N. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE*, *5*(6). doi:10.1371/journal.pone.0010937
4. Castañer, M., Torrents, C., Anguera, M.T., Dinušová, M., & Jonsson, G.K. (2009). Identifying and analyzing motor skill responses in body movement and dance. *Behavior Research Methods*, 41(3), 857–867.
5. Bishop, D.T., Wright, M.J., Jackson, R.C., & Abernethy, B. (2013). Neural Bases for Anticipation Skill in Soccer: An fMRI Study. *Journal of Sport & Exercise Psychology* 35(1), 98-109.
6. Wallace, J.L., & Norton, K.I. (2014). Evolution of World Cup soccer final games 1966–2010: Game structure, speed and play patterns. *Journal of Science and Medicine in Sport*, 17(2), 223–228.
7. Memmert, D., Hüttermann, S., & Orliczek, J. (2013). Decide like Lionel Messi! The impact of regulatory focus on divergent thinking in sports. *Journal of Applied Social Psychology* 43, 2163–2167.
8. Casarrubea, M., Jonsson, G. K., Faulisi, F., Sorbera, F., Di Giovanni, G., Benigno, A., Crescimanno, G., & Magnusson, M. S. (2015). T-pattern analysis for the study of temporal structure of animal and human behavior: A comprehensive review. *Journal of Neuroscience Methods* 239, 34-46.

9. Murgia, M., Sors, F.,Muroni, A.F.,Santoro, I., Prpic, V.,Galmonte, A., & Agostini, T. (2014). Using perceptual home-training to improve anticipation skills of soccer goalkeepers. *Psychology of Sport and Exercise,* 15(6), 642-648.
10. Teixeira, L.A., Oliveira, D.L., Romano, R.G., & Correa, S.C. (2011). Leg preference and interlateral asymmetry of balance stability in soccer players. *Research Quarterly for Exercise and Sport*, 82(1), 21–27.
11. Castañer, M., Andueza, J., Sánchez-Algarra, P., & Anguera, M.T. (2012). Extending the analysis of motor skills in relation to performance and laterality. In O. Camerino, M. Castañer & M.T. Anguera (Eds.), *Mixed methods research in the movement sciences: Case studies in sport, physical education and dance* (pp. 119-145). Abingdon: Routledge.
12. Loffing, F.; Solter, F.; Hagemann, N. & Strauss, B (2015). Accuracy of Outcome Anticipation, But Not Gaze Behavior, Differs Against Left- and Right-Handed Penalties in Team-Handball Goalkeeping. *Frontiers of Psychology* 6:1820.
13. Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments & Computers,* 32(1), 93-110.
14. Casarrubea, M., Magnusson, M. S., Roy, V., Arabo, A., Sorbera, F., Santangelo, A., Faulisi, F., & Crescimanno, G. (2014). Multivariate temporal pattern analysis applied to the study of rat behavior in the elevated plus maze: Methodological and conceptual highlights. *Journal of Neuroscience Methods* 234, 116-126.
15. Sackett, G.P. (1980). Lag sequential analysis as a data reduction technique in social interaction research. In D.B. Sawin, R.C. Hawkins, L.O. Walker., & J.H. Penticuff (Eds.), *Exceptional Infant: Vol. 4. Psychosocial risks in infant-environment transactions* (pp. 300–340). New York, NY: Brunner/Mazel.
16. Barreira, D., Garganta, J., Guimarães, P., Machado, J., & Anguera, M.T. (2014). Ball recovery patterns as a performance indicator in elite soccer. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 228(1), 61–72.
17. Camerino, O., Chaverri, J., Anguera, M.T., & Jonsson, G.K. (2012). Dynamics of the game in soccer: Detection of T-patterns. *European Journal of Sport Science*, 12(3), 216–224.
18. Jonsson, G.K., Anguera, M.T., Blanco-Villaseñor, A., Losada, J.L., Hernández-Mendo, A., Ardá, T., . . . Castellano, J. (2006). Hidden patterns of play interaction in soccer using SOF-CODER. *Behavior Research Methods*, 38(3), 372–381.
19. Magnusson, M.S., Burgoon, J.K., & Casarrubea, M. (Eds.) (2016). *Discovering Hidden Temporal Patterns in Behavior and Interaction. T-Pattern Detection and Analysis with THEME™.* Neuromethods. Springer Protocols. Human Press.
20. Perea, A., Castellano, J., Alday, L., & Hernández-Mendo, A. (2012). Analysis of behaviour in sports through Polar Coordinate Analysis with MATLAB. *Quality & Quantity,* 46(4), 1249-1260.
21. Sousa, D.J., Prudente, J., Sequeira, P., López-López, J.A., y Hernández-Mendo, A. (2015). Análisis de las situaciones de juego 2vs2 en el campeonato europeo masculino de balonmano 2012: Aplicación de la técnica de coordenadas polares [Analysis of 2vs2 situations in men's european handball championship 2012: Application of polar coordinates technique]. *Cuadernos de Psicología del Deporte,* 15(1), 181-194.
22. Anguera, M.T. (2003). Observational methods (general). In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (Vol. 2, pp. 632-637). London: Sage.
23. Anguera, M.T. & Hernández-Mendo, A. (2015). Técnicas de análisis en estudios observacionales en ciencias del deporte [Data analysis techniques in observational studies in sport sciences]. *Cuadernos de Psicología del Deporte,* 15(1), 13-30.
24. Sánchez-Algarra, P. & Anguera, M.T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: Impact of recording and coding predominating perspectives. *Quality & Quantity. International Journal of Methodology,* 47(2), 1237-1257.
25. Fernández, J., Camerino, O., Anguera, M.T., & Jonsson, G.K. (2009). Identifying and analyzing the construction and effectiveness of offensive plays in basketball by using systematic observation. *Behavior Research Methods*, 41, 719–730.
26. Lapresa, D., Arana, J., Anguera, M.T., & Garzón, B. (2013). Comparative analysis of sequentiality using SDIS-GSEQ and THEME: A concrete example in soccer. *Journal of Sports Sciences,* 31(15), 1687–1695.

27. Castañer, M., Torrents, C. Anguera, M.T. & Dinušová, M. (2008, August). Identifying and analysing motor skills answers in the corporal expression and dance through OSMOS. In A.J. Spink, M.R. Ballintijn, N. Bogers, F. Grieco, L. Loijens, L. Noldus, G. Smith, and P. Zimmerman (Eds.), *Proceedings of 6th International Conference on Methods and Techniques in Behavioral Research* (pp. 158-160). Maastricht, the Netherlands: Noldus Information Technology.
28. Byrne, D. & Ragin, C.C. (Eds.) (2015). *The Sage Handbook of case-based methods*. London: Sage.
29. Edwards, D.J.A., Dattilio, F.M., & Bromley, D.B. (2004). Developing evidence-based practice: The role of case-based research. *Professional Psychology: Research and Practice*, 35(6), 589-597.
30. Sandelowski, M. (1996). One is the liviest number: the case orientation of qualitative research. *Research in Nursing & Health*, 19, 525-529.
31. Gabín, B., Camerino, O., Anguera, M.T., & Castañer, M. (2012). Lince: Multiplatform sport analysis software. *Procedia - Social and Behavioral Sciences*, 46, 4692–4694.
32. Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge: Cambridge University Press.
33. Bakeman, R., & Quera, V. (1992). SDIS: A sequential data interchange standard. *Behavior Research Methods, Instruments & Computers*, 24(4), 554–559.
34. Hernández-Mendo, A.; Castellano, J.; Camerino, O.; Jonsson, G.; Blanco-Villaseñor, A.; Lopes, A., & Anguera, M. T. (2014). Observational software, data quality control and data analysis. *Revista de Psicología del Deporte,* 23(1), 111-121.
35. Allison, P.D. & Liker, J.K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. *Psychological Bulletin,* 91(2), 393-403.
36. Bakeman, R. (1978). Untangling streams of behavior: Sequential analysis of observation data. In G.P. Sackett (Ed.), *Observing Behavior, Vol. 2: Data collection and analysis methods* (pp. 63-78). Baltimore: University of Park Press.
37. Gorospe, G. y Anguera, M.T. (2000). Modificación de la técnica clásica de coordenadas polares mediante un desarrollo distinto de la retrospectividad: Aplicación al tenis [Retrospectivity in polar coordinates analysis: application to tennis]. *Psicothema,* 12 (Supl. Nº 2), 279-282.
38. Cochran, W.G. (1954). Some methods for streghning the common χ^2 test. *Biometrics,* 10, 417-451.
39. Fry, T.R.L., Galanos, G., & Posso, A. (2014). Let's get Messi? Top-scorer productivity in the European Champions League. *Scottish Journal of Political Economy*, 61(3), 261–279.
40. Anguera, M.T. & Hernández-Mendo, A. (2014). Metodología observacional y psicología del deporte. Estado de la cuestión [Observational methodology and sports psychology: State of affairs]. *Revista de Psicología del Deporte*, 23(1), 103-109.

Posters

The Use of Infrared Thermal Imaging in Temperature Measurement and Behavioral Assessment: Effects of Thermal Windows

C.L. Thompson[1], C. Scheidel[1], S.H. Williams[2], K.E. Glander[3] and C.J. Vinyard[4]

[1]Department of Biomedical Sciences, Grand Valley State University, Allendale, USA.
thompscy@gvsu.edu
[2]Department of Biomedical Sciences, Ohio University Heritage College of Osteopathic Medicine, Athens, USA
[3]Department of Evolutionary Anthropology, Duke University, Durham, USA
[4]Department of Anatomy and Neurobiology, Northeast Ohio Medical University, Rootstown, USA

Introduction

Body temperature is a fundamental aspect of animal biology that relates to various physiological and behavioral states including movement, emotion, social interactions, and thermoregulation [1-3]. Infrared thermal imaging has gained traction as a technique to non-invasively assess skin temperature, however these measurements have only rarely been collected simultaneously with internal temperatures from wild animals [4-5]. Infrared images can measure the surface temperature of objects, yet for most mammals, the body's surface is mostly covered in fur, with isolated areas of bare skin. The relationship between internal temperature and these thermal windows (i.e., areas that may enable heat dissipation such as bare skin) is not well established, particularly given that varying ambient temperatures will impact surface temperatures as well. We assessed the impacts of subcutaneous body and ambient temperatures on surface temperatures gained from infrared imaging of furred and bare areas of skin, and tested whether these areas serve as useful measures corresponding to thermoregulatory behaviors in a wild primate, the mantled howling monkey (*Alouatta palliata*).

Methods

We collected non-invasive infrared thermal images from wild mantled howling monkeys in Costa Rica over a two-week period during the 2014 wet season. From 113 images (n=3 animals), we obtained surface temperature measurements from the densely-furred side and back (hereafter, dorsum) areas of the body (n=86) and the face (n=32), which lacks fur. We simultaneously measured ambient temperatures from a weather station (HOBO U30, Onset Computer Corporation) and subcutaneous body temperatures by capturing animals and implanting a temperature logger (iButton, Maxim Integrated Products) between their shoulder blades, following [6]. We used the difference between concurrent subcutaneous and surface temperatures as a measure of cutaneous heat loss through the skin and fur. We also collected animals' body posture, grouped into three categories: heat-conserving (i.e., animal is fully curled, no limbs extended), intermediate (i.e., some limbs away from body and/or belly is exposed), and heat-dissipating (i.e., animal's torso is fully extended and all limbs are away from body). As the tail of mantled howling monkeys contains a furless portion of exposed skin which is potentially capable of losing heat more quickly, we also measured tail orientation, noting when the tail was fully wrapped on the body (i.e., heat-conserving) or fully exposed (i.e., heat-dissipating). We used multiple linear regression to determine the influence of subcutaneous and ambient temperatures on surface temperatures of the dorsum and face. We used a Jonckheere-Terpstra test to evaluate surface temperature and cutaneous heat loss between postures. A two-sample t-test was used to compare surface temperature and heat loss between fully wrapped and fully exposed tail postures.

Results

Dorsum surface temperatures were significantly impacted by both subcutaneous ($\beta=0.36$ $p<0.001$) and ambient temperatures ($\beta=0.51$, $p<0.001$). However subcutaneous temperature did not have a significant impact on facial temperatures ($\beta=0.08$, $p=0.73$), while ambient temperature did ($\beta=0.51$, $p=0.03$). Subcutaneous temperatures explained only 20.0% of variation in facial temperatures, compared to 47.1% of variation for dorsum

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

temperatures, with ambient temperature explaining an additional 12.2% of the variation in face temperatures (32.2% total) and 14.9% of the variation in dorsum temperatures (62.6% total).

Surface temperatures and cutaneous heat loss were also important indicators of thermoregulatory behaviors. Both dorsum and face surface temperatures were higher when animals used heat dissipating postures and lower with use of heat conserving postures (dorsum: T_{JT}=1412.0, z=5.16, p<0.001; face: T_{JT}=150.5, z=2.05, p=0.041). However, dorsum measurements also revealed an association between posture and heat loss (T_{JT}=602.5, z=-2.76, p=0.006), while face measurements did not (T_{JT}=98.0, z=-0.29, p=0.77). Similarly, animals had their tail fully wrapped against their body at lower surface temperatures for both the dorsum (t_{47}=6.61, p<0.001) and face measures (t_{15}=3.66, p=0.002). However, dorsum measurements showed a difference in cutaneous heat loss between tail wrapped and unwrapped conditions (t_{47}=-3.36, p=0.002), while face measurements did not (t_{15}=-0.82, p=0.43).

Discussion and Conclusion

Surface temperatures of densely furred regions of the body were more closely linked with subcutaneous temperature than the bare regions of the face. Hence, while the temperature of bare skin may intuitively seem like a better predictor of (subcutaneous) body temperature, this was not the case for our study. Similarly, estimates of heat loss were more closely linked to thermoregulatory behaviors for dorsum measurements than face measurements, although raw temperatures of both revealed behavioral trends. These data suggest infrared imaging can provide insight into body temperature changes and heat loss patterns, but should not be viewed as an exact analogy of body temperature as animals may use exposed body parts as areas for rapid heat loss or gain. Given the consistent impact of ambient temperature on both dorsum and face measurements, taking ambient temperature into account with thermal images can improve reliability to assess temperature and behavior.

Ethical Statement

All animal procedures were approved by Grand Valley State University, NEOMED, and Duke University IACUCs. Permits to conduct research on wild animals were granted by Costa Rica's Ministerio del Ambiente y Energia and the Organization for Tropical Studies.

References

1. Kuraoka, K. & Nakamura, K. (2011). The use of nasal skin temperature measurements in studying emotion in macaque monkeys. *Physiology & Behavior,* **102**, 347-355.
2. Meerlo, P., De Boer, S., Koolhaas, J., Daan, S. & Van den Hoofdakker, R. (1996). Changes in daily rhythms of body temperature and activity after a single social defeat in rats. *Physiology & Behavior,* **59**, 735-739.
3. Terrien, J., Perret, M. & Aujard, F. (2011). Behavioral thermoregulation in mammals: a review. *Frontiers in Bioscience,* **16**, 1428-1444.
4. Mccafferty, D.J. (2007). The value of infrared thermography for research on mammals: previous applications and future directions. *Mammal Review,* **37**, 207-223.
5. McCafferty, D.J., Gilbert, C., Paterson, W., Pomeroy, P.P., Thompson, D., Currie, J.I. *et al.* (2011). Estimating metabolic heat loss in birds and mammals by combining infrared thermography with biophysical modelling. *Comparative Biochemistry and Physiology Part A,* **158**, 337-345.
6. Thompson, C.L., Williams, S.H., Glander, K.E., Teaford, M.F., Vinyard, C.J. (2014). Body temperature and thermal environment in a generalized arboreal anthropoid, wild mantled howling monkeys (*Alouatta palliata*). *American Journal of Physical Anthropology*, **154**, 1-10.

Social Use of Facial Expressions in Several Species of Small Apes (Hylobatids)

L. Scheider[1], B.M. Waller[2], L. Oña[3], A.M. Burrows[4,5] and K. Liebal[1]

[1] Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany.
scheider@zedat.fu-berlin.de, katja.liebal@fu-berlin.de
[2] Department of Psychology, University of Portsmouth, Portsmouth, United Kingdom.
Bridget.Waller@port.ac.uk
[3] Department of Natural Science and Mathematics, University of Groningen, Groningen, The Netherlands.
l.ona@rug.nl
[4] Department of Physical Therapy, Duquesne University, Pittsburgh, Pennsylvania, United States of America.
burrows@duq.edu
[5] Department. of Anthropology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

Introduction

Intentional use is a key feature of human language. Therefore, in searching for the evolutionary roots of human communication, comparative researchers have dedicated much attention to the question whether communication of non-human primates is also characterized by their intentional use of different signal types. Thus, they should use their signals purposefully and direct them to other group members, indicating that they have some voluntary control over the production of their signals. Non-human primates use various communicate means in interactions with others. The current status quo in research into non-human primate communication is that gestures are commonly considered intentionally produced and flexibly used signals, while vocalizations and facial expressions are often described as involuntary expressions of affective internal states [1]. Only a few studies investigated how the use of facial expressions in nonhuman primates is adjusted to the recipient's behavior [2,3].

Method

In order to explore how non-human primates use facial expressions in specific communicative interactions, we studied five species of small apes by using a Facial Action Coding System for *Hylobatids* (GibbonFACS) [4]. Most existing research on facial expressions has been devoted to investigating the role of facial expressions for coordinating social interaction, facilitating group cohesion and the maintenance of individual social relationships in non-human primates, including several species of monkeys and great apes (e.g., *Pan troglodytes*, *Pongo pygmaeus*, *Macaca mulatta*, *Callithrix jacchus*), but little is known about the various species of small apes or gibbons (Hylobatidae). Gibbons are equipped with extensive facial muscles [5], which they use to perform a variety of facial movements [6]. The newly developed GibbonFACS offers the opportunity to examine facial expressions in much greater detail than previous methods and enables the objective and standardized comparison of facial movements across species. For the development of such a coding system, the muscular movements underlying facial expressions are identified and used to define facial expressions as a combination of such facial muscle contractions (Action Units [AUs]) or more general head/eye movements (Action Descriptors [ADs]). If gibbons are capable of adjusting the use of facial expressions to the behaviour of the recipient and the context in which they are used, we predict that 1) the sender uses facial expressions with longer duration when the recipient is facing the sender than when they are not, 2) the sender uses facial expressions more frequently in social than in non-social contexts when the recipient was facing the sender, and 3) the events of consecutive facial expressions, indicating that the recipient responds to the sender's facial expression by producing another facial expression is more common when individuals are facing each other than when they are not.

Results

We found that the duration of facial expressions was significantly longer when gibbons were facing their pair partner compared to non-facing situations. Additionally, gibbons used facial expressions while facing more often in social contexts than non-social contexts where facial expressions were produced regardless of the attentional

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

stance of the partner. Also, facial expressions were more likely ‘responded to’ by the partner’s facial expressions when facing than non-facing. Taken together, our results indicate that gibbons use their facial expressions depending on their social environment and use them in a directed way in communicative interactions with conspecifics.

This is an observational, non-invasive study. All procedures were performed in full accordance with German legal regulations and the guidelines for the treatments of animals in behavioural research and teaching of the Association for the Study of Animal Behaviour (ASAB).

Keywords: communication, gibbons, GibbonFACS, facial expression

References

1. Tomasello, M. (2008). The Origins of Human Communication. London. The MIT Press.
2. Waller, B.M., Caeiro, C.C., Davila-Ross, M. (2015). Orangutans modify facial displays depending on recipient attention. *PeerJ* **3**: e827.doi:10.7717/peerj.827.
3. Liebal, K., Pika, S., Tomasello, M. (2004). Social communication in siamangs (*Symphalangus syndactylus*): use of gestures and facial expressions. *Primates* **45**, 41–57.
4. Waller, B.M., Lembeck, M., Kuchenbuch, P., Burrows, A.M., Liebal, K. (2012) GibbonFACS: A Muscle-Based Facial Movement Coding System for Hylobatids. *International Journal of Primatology*.**33**, 809–821.
5. Burrows, A.M., Diogo, R., Waller, B.M., Bonar, C.J., Liebal, K. (2011). Evolution of the muscles of facial expression in a monogamous ape: evaluating the relative influences of ecological and phylogenetic factors in hylobatids. *The Anatomical Record* **294**, 645-63.
6. Scheider, L., Liebal, K., Oña, L., Burrows, A., Waller, B. (2014). A comparison of facial expression properties in five hylobatid species. *American Journal of Primatology* **76**, 618–628.

Sensor Enabled Affective Computing for Enhancing Medical Care (SenseCare)

A. Keary[1], F. Engel[2], R. Bond[3], M.D. Mulvenna[3], M. Hemmje[2], HY. Wang[3], H. Zheng[3], R. Riestra[4] and P. Walsh[1]

1 Department of Computing, Cork Institute of Technology, Cork City, Ireland. {alphonsus.keary, paul.walsh}@mycit.ie
2 FTK – Research Institute for Telecommunication and Cooperation, Dortmund, Germany. fengel@ftk.de, matthias.hemmje@fernuni-hagen.de
3 Ulster University, Newtownabbey, Northern Ireland. {rb.bond, md.mulvenna, hy.wang,h.zheng}@ulster.ac.uk
4 Groupo INMARK, Madrid, Spain. ruben.riestra@grupoinmark.com

Introduction

SenseCare (http://www.sensecare.eu/) is an affective computing (Picard, 1997) and sensory fusion based medical and connected healthcare project being led by the Cork Institute of Technology (CIT). SenseCare brings together a diverse group of subject matter experts from industry and academia with the main aim of enhancing and advancing future connected healthcare/eHealth processes and systems using sensory and machine learning technologies to provide emotional (affective) and cognitive insights into patients well-being so as to provide them with more effective treatment across multiple medical domains. The main objective of the SenseCare project is to develop technologies and methods that will lessen the enormous and growing health care costs of dementia and related cognitive impairments that burden European citizens, which is estimated to cost over €250 billion by 2030 (Wimo, Jönsson, & Gustavsson, 2009) (Wimo, Jönsson, & Gustavsson, 2009). Our presentation provides an overview of the proposed SenseCare platform along with a discussion on the main aims and objectives of the project. An overview of the platform architecture and the core technical aspects of the system will also be presented.

Objectives of SenseCare

The primary objective of the SenseCare project is to "develop a cloud based affective computing operating system capable of processing and fusing multiple sensory data streams to provide cognitive and emotional intelligence for connected healthcare systems". In particular the SenseCare consortium will:

- Specify and engineer the architecture of the SenseCare platform and release two versions of the platform infrastructure during the life of the project (2016 – 2019).
- Create and evaluate two use case test pilots (relating to the dementia care and connected health domains) that integrate with, use and apply the services of the SenseCare platform.
- Specify and engineer a number of medical informatics applications that will run on the SenseCare platform and that will also be tested and evaluated as part of the use case test pilot phases.

SenseCare Platform Overview

Initially SenseCare software services will be applied to the dementia care and connected health domain where enormous potential and opportunities exist in relation to providing intelligence and assistance for self-managed care possibilities to people with dementia, co-managed care alongside family member and other informal carers, as well as offering insights on emotional state and overall wellbeing to healthcare professionals. Figure 1 provides our interpretation of a typical use case scenario for the SenseCare platform. SenseCare is envisioned as a platform that enables the integration, fusion, analysis and archival of various sensor data streams to provide a holistic insight/view on the wellbeing of a person with dementia as well as comprehensive access for later information reuse. The platform will be engineered using the "*Educational Portal (EP) tool suite that was* *developed by the software company GLOBIT. The EP tool suite was developed as a solution for the growing needs of scientific communities to manage their documents and media collections as well as their educational and other kinds of knowledge resources*" (Nawroth, et al.).

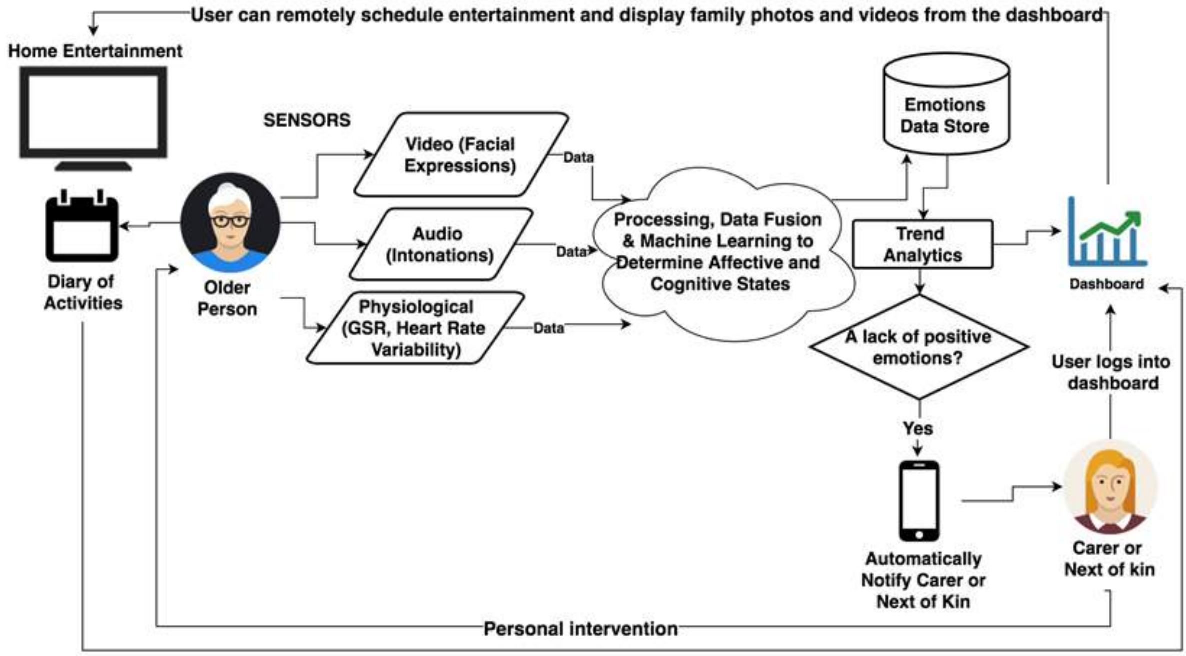


Figure 1: Framework for monitoring and caring for the emotional wellbeing of older people.

The EP is a flexible software solution, that is built around the content management functionality that is provided by the Typo3 framework (TYPO3.ORG, 2016). The modular setup of the Typo3 architecture allows for the easy extension of functionality through the provision of plugin functionality. Several of these plugin extensions, that actively support the pre-processing of various kind of research data already exist in the EP and have been successfully applied and integrated into various research projects conducted by the SenseCare consortium.

To suit the functionality requirements specified in SenseCare, the EP will be harmonised and extended. To address this, we envision the addition of semantic representation of all processes of the treatment and care of the people with dementia (Hallberg, et al., 2007). The EP infrastructure will be used as an anchor for all the digital assets (content and knowledge resources) that need to be kept in the context of the processes associated with the wellbeing and healthcare of the person with dementia. For example, the archival of videos will apply the MPEG7 standard that supports the semantic enrichment of videos for later reuse. Furthermore, this representation of workflows and involved resources will be aligned with an additional schema that will be applied in order to integrate semantically or fuse the various incoming affective and physiological data streams. This contextualisation of managed resources ensures on the one hand a rich and consistent data source to visualise the emotional state of the person with dementia and all involved resources, besides a consistent entry point for retrieving provided data for its further comprehensive reuse. To support these undertaking, we expect to provide interfaces for various kinds of sensor data besides services for sensor data feature extraction, data filtering, quality assurance and retrieval. The section below provides a deeper description on the type of data streams that we envision integrating into the SenseCare platform.

Emotion signals (data streams)

Video based emotional data streams: Emotient provide an API that uses video based frame-by-frame analysis of emotional responses. The Emotient platform provides data on 19 Facial Action Units, which are the elementary facial muscle movements based on Dr. Paul Ekman's Facial Action Coding System (FACS) (Paul Ekman Group, 2015), (Ekman & Friesen, 1978). The Intel Real Sense platform (Intel, 2014) is also another vision based SDK platform that offers raw vision based data feeds that can be used and analysed by the SenseCare platform. Affectiva (Affectiva, 2015) also provide vision based emotion analytics and claims to have processed almost four million facial images using their AI based platform. SenseCare will be designed to interface with vision based emotional analytical platforms similar to those discussed above.

Wearable based emotional data streams: Shimmer is an established provider of sensory physiological computing for affect recognition using GSR (Galvanic Skin Response), ECG (ElectroCardioGram) and EMG (ElectroMyoGram) (Shimmer, 2015). Empatica has its origins in the affective sciences and specialises in clinical grade wearables. Their research focused E4 wrist worn device provides embedded sensors for photoplethysmography (PPG for continuous hearth rate analytics), electrodermal activity (EDA analytics for galvanic skin response, sympathetic activation, autonomic arousal and excitement), a 3-axix accelerometer for movement data and an infrared skin temperature thermopile. The Empatica Connect cloud service provides access to encrypted data, CSV format downloads, precise time stamped data and signal comparison data services (Empatica, 2015). SenseCare architecture will also be designed to interface with wearables and IoTs based emotional analytical services and platforms similar to those described above.

Other sources of emotional data streams: There are ever increasing sources of affective data streams emerging from academic laboratories and reaching real-world application status. Voice emotion analytics (Schuller , Eyben, & Weninger, 2016) & (Viszlay, Juhár, & Pleva, 2012), EEG signal processing (Emotiv, 2015), affective gait analytics (PSFK.COM, n.d.) are also expected to be future sensory fusion interface candidates for inclusion in further iterations of the SenseCare platform.

Visualization of emotion signals

An interactive visualization model will also be implemented to allow users to interrogate the various kind of streaming data, described above at 'different levels of detail'. The highest level will most likely incorporate radar plots to visualize the frequency of emotions with reference to two baselines (a personal baseline [the normal frequency of each emotion for that individual] and a peer baseline [the mean frequency of each emotion as derived from all participants using the SenseCare platform]). A colour scheme similar to a traffic-lights approach will be taken to provide an effective visual hierarchy (making critical information unavoidable to the human visual system). Temporal visualisations will be used to depict the changes in emotions over time, which will indicate days and times when occupants are most vulnerable, as well as indicating any correlations or patterns between affective states and activities (e.g. visitations and TV episodes etc.). Spatial visualisations will also overlay patterns in emotions and the participants' geo-locations. The novelty will involve the semantic search-ability of the data in the SenseCare dashboard and the ability for data exploration and interrogation at multiple levels. All visualisations will be co-created with users and will be underpinned by principles taken from cognitive and Gestalt psychology as well as human factors research.

Along with the presentation of the platform, affective sensory technologies and visualisation techniques there will also be an insight into a number of SenseCare use case scenarios that have been developed by the consortium to date.

Acknowledgements

This publication has been produced in the context of the SenseCare project. This project has received funding from the European Union's H2020 Programme under grant agreement No 690862. However, this paper reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

References

1. Affectiva. (2015, July 06). Affdex. Retrieved July 06, 2015, from http://www.affectiva.com/solutions/affdex/
2. Affectiva. (2015). *Affdex*. Retrieved May 07, 2015, from http://www.affectiva.com/affdex/#pane_overview
3. Ekman, P., & Friesen, W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Palo Alto California: Consulting Psychologists Press.
4. Emotient. (2013, December 12). *Emotient Software Development Kit*. Retrieved July 07, 2015, from http://www.emotient.com/about/press/emotient-launches-new-software-development-kit-for-real-time-emotion-recognition/
5. Emotient. (2014, February 27). *Emotient Announces Series B Financing Of $6M*. Retrieved July 02, 2015, from http://www.emotient.com/about/press/emotient-announces-series-b-financing-of-6m/
6. Emotiv. (2015, 02 05). Emotiv *EPOC*. Retrieved 02 05, 2015, from http://emotiv.com/
7. Empatica. (2015, July 02). *Empatica E4 Wristband*. Retrieved July 02, 2015, from https://www.empatica.com/e4-wristband
8. Hallberg, J., Nugent, C. D., Davies, R. J., Synnes, K., Donnelly, M. P., Finlay, D., & Mulvenna, M. (2007). HomeRuleML - A Model for the Exchange of Decision Support Rules Within Smart Environments. *IEEE International Conference on Automation Science and Engineering (CASE)* (pp. 513 - 520). Scottsdale, AZ: IEEE. doi:10.1109/COASE.2007.4341727
9. Intel. (2014). *Real Sense SDK*. Retrieved July 07, 2015, from https://software.intel.com/sites/default/files/managed/0f/b0/IntelRealSense-WindowsSDKGold_PB_1114-FINAL.pdf
10. Nawroth, C., Schmedding, M., Brocks, H., Kaufmann, M., Fuchs, M., & Hemmje, M. (n.d.). Towards Cloud-Based Knowledge Capturing Based on Natural Language Processing. *Procedia Computer Science, 68*, 206-216.
11. Paul Ekman Group. (2015, July 18). *Paul Ekman*. Retrieved July 18, 2015, from http://www.paulekman.com/
12. Picard, R. (1997). *Affective* Computing. MIT Press Cambridge, MA, USA ©1997.
13. PSFK.COM. (n.d.). *Kinect system reads emotions to help autistic children socialize*. Retrieved 03 14, 2016, from http://www.psfk.com/2014/03/kinect-autism-emotion-detection.html#!D1LyT
14. Schuller , B., Eyben, F., & Weninger, F. (2016, 03 09). *OpenAudio*. Retrieved 03 09, 2016, from http://www.openaudio.eu/
15. Shimmer. (2015). Shimmer *Sensing Technologies*. (Shimmer) Retrieved April 17, 2015, from http://www.shimmersensing.com/
16. TYPO3.ORG. (2016, 03 21). *TYPO3*. (TYPO3.ORG) Retrieved 03 21, 2016, from https://typo3.org/about/
17. Viszlay, P., Juhár, J., & Pleva, M. (2012). Alternative Phonetic Class Definition in Linear Discriminant Analysis of Speech. *19th IEEE* International *Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 655-658). Vienna: IEEE. Retrieved from http://www.kemt.fei.tuke.sk/~pleva/down/papers/1569528635%20iwssip%202012%20vienna.pdf
18. Wimo, A., Jönsson, L., & Gustavsson, A. (2009, 10 27). *Cost of illness and burden of dementia in Europe - Prognosis to 2030.* (A. Europe, Ed.) Retrieved 04 23, 2015, from http://www.alzheimer-europe.org/Research/European-Collaboration-on-Dementia/Cost-of-dementia/Prognosis-to-2030

Uncovering Aspects of Teacher Professional Identity

Graham J. Passmore

Education Department, Lakehead University, Thunder Bay, Canada
gpassmore@lakeheadu.ca

Introduction

In a review of teacher identity research Beijaard, Meijer and Verloop (2004) reported that Knowles (1992), and Nias (1989), related professional identity to images of self while Goodson and Cole (1994), Volkmann and Anderson (1998), Cooper and Olson (1996) placed emphasis on teacher roles. In regard to roles Beijaard et al (2004) suggested that four characteristics were of import: commitment to teaching, professional orientation, task orientation, and self-efficacy. On applying scales to measure these characteristics Lamote and Engels (2010) and Rots, Aelterman, Vlerick, and Vermeulen (2007) successfully distinguished among higher, median and lower ratings. These workers noted that greater differentiation among teacher identity requires a scale that is sensitive to processes and factors and that influence identity formation such as: the continuous evaluation and re-evaluation of experiences (Kerby, 1991), social context and concepts of self (Beijaard, Meijeer and Verloop, 2004), personal history and culture (Lamote and Engels, 2010).

Most identity assessment methods (Duriez and Soenens 2006; Cokley 2007; Crocetti, Rubini, Berzonsky and Meeus 2009) assemble existing instruments to address a specific question of identity and are based in a single theory of identity (Passmore, Ellis and Hogard, 2014). Some (Akhtar and Samuel, 1996; Berzonsky; 1999) introduce new psychometric measures but, similar to the ad hoc approach, they are designed to address a specific question and cannot be adapted to questions of teacher identity. In contrast, the ISA (Identity Structure Analysis) method is based in a meta-theoretical framework and is amenable to the development of instruments that are sensitive to knowledge of self and personal and contextual facets of identity (Weinreich and Saunderson, 2003, Hogard, 2014). This work presents a new way to organize and present results of analyses of the 20 page reports that the ISA method generates. In doing so it demonstrates differentiation among the professional identities of 2 teachers and indicates that the ISA meets the call for sensitivity to the factors and processes of identity formation.

Method

An ISA instrument was developed to assess teacher identity. Like all ISA instruments it contains bipolar constructs and entities (Appendix 1). The constructs aligned to themes (team player, class management, relationship with students, approach to work and problem solving) within the topic under consideration (teaching) and their poles were sensitive to tensions within the themes. Entities in ISA are people, institutions or icons in salient life domains (self, work, home and society) that influence identity formation, The fact that constructs and entities of an ISA instrument align to themes and domains is the reason that the ISA method is sensitive to knowledge of self, and personal and contextual facets of identity. A dedicated ISA software (Ipseus) generates a matrix of constructs and entities on the fly and at random for each person sitting an instrument. The combinations of construct and entity that make up the matrix are presented one at a time and in turn as per Figure 1.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

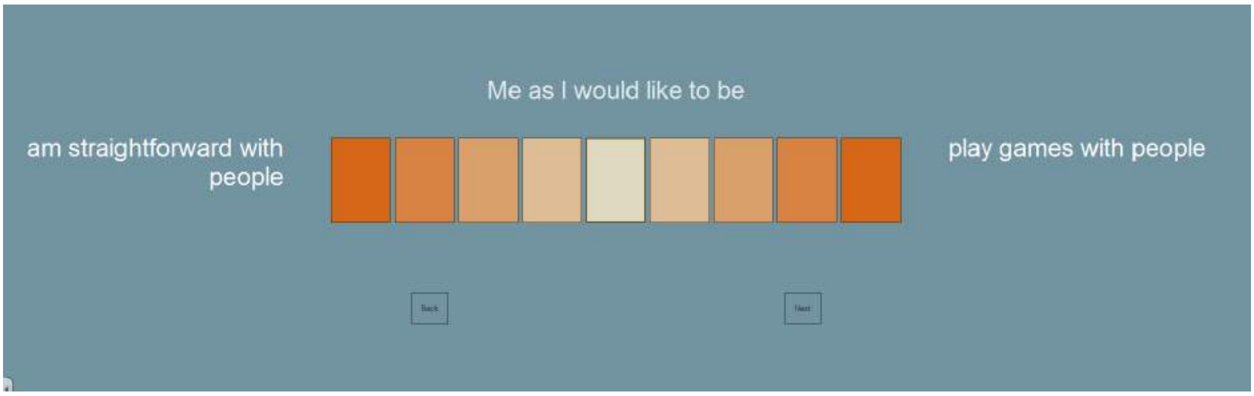


Figure 1. Presentation of the instrument matrix in Ipseus

The first construct is rated along the 9-point zero-center scale in the guise of the entity of that particular instance of the matrix. The second construct-entity combination is then presented for rating. The process continues until attributions are made for every construct/entity combination in the matrix. The nature of entities and bipolar constructs means that ratings along the zero-center scale reveal a person's values and beliefs for aspects of self and influential others. The Ipseus software converts the inputted ratings to standard scores, and enters these into a series of formulae. The formulae are algebraic mirror images of ISA identity parameters (definitions of the parameters and their algebraic formulae are available for review in chapters 1 and 2 of Weinreich and Saunderson (2003)). The use of standard scores means that comparison of ISA parameter values is possible across groups and across individual teachers. The parameter values are made available as a 20page ISA/Ipseus report. Recognizing that interpretation of a 20-page ISA/Ipseus report is a considerable undertaking Weinreich (2013) developed a template to streamline the interpretation process.

In summary the template calls for review of the ISA parameters in the following order

- Report core and conflicted dimensions of identity.
- Report idealistic and contra-identifications with influential others.
- Compare empathetic identifications for 2 current entities of self and one past entity of self.
- Report identity conflicts for 2 current entities of self and one past entity of self.
- Report on evaluations of and ego-involvement with influential others.
- Report on the intersection of evaluations of entities of self and identity diffusion.

In this work I have adapted the report template to consider the professional identities of 2 teachers so as to indicate that discrimination among individual teachers' professional identity is possible at a granularity that exceeds high, medium or low. The teacher data is taken from an ongoing (2015-2016 academic year) study involving Hong Kong teachers. The teachers complete the teacher ISA instrument using a Google form. The form data is downloaded and imported into Ipseus for analysis. Beyond completing the ISA instrument the teachers provide demographic information: gender, subject taught, and years in teaching. A rating of the teachers conducted by their principal (good, average or weak) serves as an additional source of teacher information. The rubric that the principals use to rate the teachers is presented in Appendix 2. The Hong Kong teachers considered are labelled Teacher A (a 'good' male teacher of English with 4 years of experience), and Teacher B (a female teacher of Business with 6 years of experience and rated as weak to average be her principal).

Core and Conflicted Evaluative Dimensions of Identity

Structural pressure (SP) is an ISA parameter that reflects the consistency with which a construct is used to evaluate entities. High structural pressure is associated with constructs that are used in consistent manner to evaluate others. These constructs represent the core, stable evaluative dimensions of the identity under consideration. Low structural pressure is associated with constructs that are used to evaluate others in different ways depending on circumstance and context. Low structural pressure suggests an area of stress and indecision; a conflicted area liable to poor decision making. Core and conflicted constructs for both teachers are presented below.

Table 1. Structural Pressures for Teacher A

Core Constructs		
Pole 1	Pole 2	
believes there is no finer job than teaching	believes there are better jobs than teaching	89.36
sides with society's disadvantaged	sides with the advantaged in society	77.72
feels there is a lot I can do to get students to value learning	feel there is little I can do to get students to value learning	77.47
follows a firm agenda when dealing with difficulties	*deals with difficulties creatively*	72.09
deals with awkward people by appealing to everyday rules	*confronts awkward people*	71.46
prioritizes achievement	*prioritizes welfare*	58.86
Conflicted Constructs		
puts the needs of students first	puts personal needs first	11.26
becomes closely involved with students	is remote from students	-7.01
Is straightforward with people	*plays games with people*	-37.55

This teacher expresses strong commitment to the job of teaching by way of a strong endorsement of the statement that 'there is no finer job than teaching.' Importantly he feels there is a lot he can do to get students to value learning and he is oriented toward serving the disadvantaged in society. In short this is a teacher who encourages students in their academic endeavours. In regard to class management Teacher A will stand up to awkward individuals but the potential for confrontation with tough students will likely be offset though by his orientation toward creative problem solving.

Areas where Teacher A experiences stress and a measure of vacillation include placing the needs of students over the needs of self and the related matter of trying to find a balance between his involvement with students and the need to maintain a degree of distance from them. The mix of personal and professional that the job demands requires that the successful teacher walk a fine line. This teacher is well versed in these difficulties. Interestingly Teacher A expressed confliction over the construct playing games with people as opposed to being a straight shooter. A follow up interview is recommended to determine the degree to which this finding reflects the expressed stance toward creative problem solving and class management.

Table 2. Structural Pressures for Teacher B

Core Constructs		SP
Pole 1	Pole 2	
communicates well with parents	is remote from parents	100.00
feels there is a lot I can do to get students to value learning	feel there is little I can do to get students to value learning	81.48
puts the needs of students first	puts personal needs first	61.38
is straightforward with people	plays games with people	53.97
Conflicted Constructs		
believes there is no finer job than teaching	believes there are better jobs than teaching	14.81
becomes closely involved with students	is remote from students	11.11
Prioritizes achievement	Prioritizes welfare	3.17

Table 2 points to a teacher who places greatest emphasis on effective communication with parents. Perhaps these communications are seen as a route to improved class management. On the plus side she feels she can get students to value learning and she expresses a desire to work straightforwardly with others. Placing the needs of students first is an admirable quality but mentoring may be required to offset the potential this construct holds for influencing work/life balance. Conflicted constructs point to areas where additional mentoring would be of value. In particular, vacillation over the notion that there are better jobs than teaching is worthy of attention. It points to potential for attrition from the profession. On the class management front this teacher struggles with the issues of becoming involved with her students and caring for their welfare. Mentoring could help this caring teacher see the need to be somewhat remote from students so as to better cater to their academic achievement and her own needs.

Idealistic and contra-identifications with others

Idealistic identifications (II) point to a person's role models. They indicate the characteristics a person will seek to emulate over the long term. Contra-identifications (CI) indicate negative role models. Those who possess characteristics from which a person wishes to dissociate

Table 3. Idealistic Identifications for Teachers A and B

Teacher A	**II**	**Teacher B**	**II**
Experts in literature	0.83	A person I admire	0.67
A good teacher	0.50	Closest family member	0.67
A good student	0.50	A good teacher	0.58
A disruptive student	0.50	School Principal	0.58
	CI		**CI**
A person I do not like	0.75	A person I do not like	0.67
Experts in government…	0.67	A disruptive student	0.42

Teacher A shows a strong positive orientation toward the subject matter he teaches. His aspirations toward functioning as a good teacher and as a good student are moderately high. Interestingly, this 'good' teacher also expresses a desire to function as a disruptive student. The contra-identifications of Teacher A are unremarkable. Teacher B's primary behavioural orientations are toward a person she admires and her closest family member. Interestingly she exhibits contra-identification with a disruptive student. Mentoring to uncover the source of this conflict might help her with class management issues and to come to an improved stance toward the job of teaching.

Empathetic identifications with others

Idealistic identifications represent long-term aspirations; empathetic identifications are of the here and now. Change in empathetic identifications across context and mood states reflect potential for change in behaviour. For teacher A the strongest empathetic identification at work is with experts in literature (0.7), the school principal (0.7) and a good teacher (0.5). At home the identification pattern with a good teacher is maintained. Interestingly, empathetic identification with the school principal drops off (0.56) while association with a disruptive student (0.3-0.44) increases. Overtime in the work environment empathetic identification with a good student has declined (0.55-0.4) and empathy with a disruptive student has increased (0.27-0.3).

Regardless of mood, context and stage of biography Teacher B's most extensive empathetic identifications are with: a person I admire, my closest family member and the school principal. Empathetic identification with a disruptive student is low regardless of mood, context and stage of biography. Of particular note is the fact that this teacher's identification with entities that represent the subject matter she teaches has declined over time both at work and at home.

Conflicted Identifications with others

Conflicted identification in ISA references being ‘as’ another while at the same time wishing to disassociate with the characteristics that are seen to be held in common. For teacher A, notable conflicted identification occurs with ‘experts in government, social policy, human rights and health’ at home and at work. For Teacher B conflicted identification is seen for ‘the school principal’ both at work and at home. Conflicted identification with her principal could be due to the high level of empathetic identification this teacher has for this entity; questioning to confirm that this is the case would be beneficial.

Evaluation of, and ego-involvement with, entities of primary investigative interest

For teacher A the entities of interest are: experts in literature, good teacher, good student and the disruptive student. Evaluation (a parameter that ranges from- 1.00 - +1.00) of these entities respectively are 0.87, 0.67, 0.00 and 0.1. Looking at ego involvement (0.00-5.00) with these entities, the report indicates that experts in literature and the disruptive student hold more influence over this teacher (3.41 and 3.17 respectively) than the good teacher (1.59) and the good student (1.71). It is worth noting that for this teacher ego-involvement with the good teacher and the good student is rather low suggesting limited motivation to behave like them.

For teacher B the entities of primary investigative interest are: a person I admire, my closest family member, a good teacher, the school principal and a disruptive student. Ego involvement with these entities (4.17, 4.58, 2.92, 5.00 and 2.92) ranges from moderately high to the maximum. The influence that the school principal holds over this teacher is particularly worthy of note. Evaluations of the first 4 of these entities (0.91, 0.51, 0.73, 0.44) are positive while the disruptive student (-0.61) garners a negative evaluation. A point of note is that the negative evaluation of the disruptive student is accompanied by moderately high ego-involvement. Again, mentoring to help this teacher reach disruptive students would be of use. Mentoring to limit the principal’s influence is also recommended.

Evaluation of self and Extent of Identity Diffusion

In the case of teacher A, the ISA report points to an individual of moderately high levels of diffusion (low diffusion being 0.2 or less and moderately high diffusion being any value greater than 0.4) both at work (0.42) and in the guise of the entity ‘me as a student teacher’ (0.43). Fortunately, Ego-involvement with these entities is (1.95 and 2.32 respectively) low indicating that the sense of diffusion experienced is not troubling. Teacher B exhibits low levels of diffusion across all entities of self. On the other hand her ego-involvement with all entities of self is high and in particular with the entity ‘me at work’ (4.58). In other words, Teacher B is far more influenced by how she sees herself performing than is the case for teacher A. Both teachers exhibit extreme negative evaluations and high levels of ego-involvement with the entity ‘me as I would hate to be.’ Both teachers are driven to an extreme degree to avoid being “as” this entity of self.

Identity Variants

The ISA report presents a table of entities of self that plots evaluation against identity diffusion. The table is

Defensive High Self-Regard	Confident	Diffuse High Self-Regard
Defensive	Indeterminate	Diffusion
Defensive Negative	Negative	Crisis

Figure 2. Demarcation table of evaluation vs diffusion

divided into a series of sections. A snapshot of the status of an identity is available from the sections where the entities of self lie. In the case of teacher A, most entities of self are associated with moderate or reasonable levels of self-evaluation. The entities 'me at work' and me as a student teacher' are associated with moderately high levels of diffusion such that they fall where the ideal 'indeterminate section meets the diffusion section. Lower diffusion for the entity 'me at home' places this entity of self in the desired indeterminate region. A higher level of self-evaluation is associated with the entity "me as I would like to be" indicating that this teacher accepts that there is room for improvement. An interview to determine how this teacher aims to work to make this "me as I would like to be" a reality would be helpful.

In the case of Teacher B most all of the entities of self fall in the quadrant that ISA labels defensive high self-regard (unreasonably high self-evaluation and low diffusion). While on the outside this teacher may appear to be fine, pointing her to better perspectives and conclusions about her teaching may come as a heavy blow.

Concluding Remarks

Teacher A was rated as a good teacher of English by his principal. He exhibits a high level of commitment to the job of teaching and a love for his subject matter. Built into the core of the professional identity of this teacher is an approach to classroom management that sees him confronting awkward people and working toward student achievement. He struggles with maintaining work/life balance in regard to getting close to his students. Vs maintaining a formal distance. Teacher B was rated as an average teacher of Business Management. The prominent core construct of her identity is her orientation toward positive communication with parents. This finding may be a compensatory mechanism to overcome noted vacillations in regard to class management constructs. Teacher B appears to lack commitment to the job of teaching. Part of her troubles lie with the entity 'a disruptive student.' Mentoring to help her realise greater empathetic identification with this entity may help her with class management issues.

Noted prior research into teacher identity Beijaard, Meijer and Verloop (2004), Lamote and Engels (2010) and Rots (2007) elicited gross differentiations among teachers. The very same work notes that more granular differentiation requires a scale that is sensitive to the continuous nature of identity formation (Kerby, 1991), social context and concepts of self (Beijaard, Meijeer and Verloop, 2004) and personal history and culture (Lamote and Engels). This work has illustrated that ISA meets the exhortations of earlier researchers by way of the constructs and entities that make up its instruments. Entities include salient people, institutions and social icons that influence identity formation in regard to culture and personal history. Entities of self (current self in various life domains, past and future entities of self) accommodate biographical stages. Bipolar constructs render ISA sensitive to issues acting within life domains upon which a person will ruminate as identity formation occurs. Further this work has illustrated that the new ISA interpretation template is capable of ready interpretation of the 20-page ISA/Ipseus report and the generation of comparisons of individual identities.

References

1. Akhtar and Samuel (1996). The concept of identity: Developmental origins, phenomenology, clinical relevance and measurement. *Harvard Review of* Psychiatry, Jan-Feb; 3 (5):254-67
2. Beijaard, D., Meijer, P.C., Verloop, N. (2004). Reconsidering research on teacher identity, *Teaching and Teacher Education* 20 (2)1070128
3. Berzonsky, M.D. (1999) Identity styles and hypothesis testing strategies. *Journal of Social Psychology*, 139(6) 78-789
4. Cokley, K. (2007). Critical Issues in the Measurement of Ethnic and Racial Identity: A Referendum on the State of the Field. *Journal of Counselling Psychology*, 54 (3), 224-234
5. Cooper, K., & Olson, M. R. (1996). The multiple 'I's' of teacher identity. In M. Kompf, W. R. Bond, D. Dworet, & R. T. Boak (Eds.), *Changing research and practice: Teachers 'professionalism, identities and knowledge* (pp. 78–89).London/Washington, DC: The Falmer Press.
6. Crocetti, E., Rubini, M., Berzonsky, M. D., & Meeus, W. (2009). Brief Report: The Identity Style Inventory: Validation in Italian adolescents and college students. *Journal of Adolescence*, 32, 425-433..
7. Duriez, B. Soenens, B. (20070). Personality, identity styles, and religiosity: an integrative study among late and middle adolescents. *Journal of Adolescence*, Feb 29 (1), 119-35
8. Hogard, E. (2014). Evaluating the Identity of Program Recipients Using and Identity Exploration Instrument. *Canadian Journal of Program Evaluation*, (spring) 1-35
9. Kerby, A. (1991). Narrative and the self. Bloomington, IN: Indiana University Press.
10. Knowles, G. J. (1992). Models for understanding pre-service and beginning teachers' biographies: Illustrations from case studies. In I. F. Goodson (Ed.), *Studying teachers' lives* (pp. 99–152). London: Routledge.
11. Lamote, C., & Engels, N. (2010). The development of student teachers' professional identity. *European Journal of Teacher Education,* 33(1), 3-18.
12. Nias, J. (1989). Teaching and the self. In M. L. Holly, & C. S. McLoughlin (Eds.), *Perspective on teacher professional development* (pp. 151–171). London: Falmer Press
13. Passmore, G., Ellis, R., and Hogard, E., (2014) Measuring Identity: A Review of Published Studies Buckinghamshire New University *SHEU Occasional Paper* Number 121.
14. Rots, I., Aelterman, Vlerick, P., and Vermeulen, K. (2007). Teacher education, graduates' teaching commitment and entrance into the teaching profession. Teaching *and teacher Education*, 23, 543-556
15. Weinreich, P., & Saunderson, W. (Eds.) (2003/2012) *Analysing Identity: Cross-Cultural, Societal and Clinical Contexts.* London & New York: Taylor & Francis/Routledge/Psychology Press.
16. Weinreich, P. (2013). Identity Structure Analysis template for results interpretation 2013 <https://drive.google.com/file/d/0B3x9x0GcOogoZzFIei1qelFNUEk/view >Accessed Jan 27, 2015

Appendix 1. The Teacher Identity Instrument

The 12 Constructs

Theme 1: Team Player

Is straightforward with people… Plays games with people

Takes issue with the way things are…Supports the way things are

Communicates well with parents… Is remote from parents

Theme 2: Approach to Class Management and Teaching

Prioritizes achievement….Prioritizes welfare

feel there is a lot I can do to get students to value learning…. feels there is little I can do to get students to value learning

deals with awkward people by appealing to every day rules…confronts awkward people

Theme 3: Relationship with Students

sides with society's disadvantaged …sides with the advantaged in society

puts the needs of students first…puts personal needs first

becomes closely involved with students…maintains a formal relationship

Theme 4: Approach to Work and Problem Solving

depends on others in making decisions… prefers to work things out alone

follows a firm agenda when dealing with difficulties….deals with difficulties creatively

believes there is no finer job than teaching…believes there are better jobs than teaching

The 15 Entities

Required Entities of Self:

Me as I would like to be

Me as I would hate to be

Me at home

Me at work

Me as I used to be

Domain 1: Home

My closest family member

Domain 2: Work

A good teacher

A good student

A disruptive student

School Principal

Domain 3: Society

Experts in mathematics science and technology

Experts in literature, art, music, dance, film, media and history.

Experts in government, social policy, human rights and health

A person I admire

A person I do not like

Apppendix 2

Teacher Quality Rubric

Question	*Weak*	*Average*	*Strong*
Teacher Efficacy			
To what extent does this teacher make use of a variety of instructional strategies			
To what extent does this teacher gauge student comprehension and provide alternate explanations?			
To what extent does this teacher adjust assessment strategies for your students?			
To what extent does this teacher get all students to believe they can do well in school and value learning?			
Classroom Management			
To what extent does this teacher control disruptive behaviour and get students to follow class rules?			
Professional Orientation			
To what extent does this teacher keep abreast of educational findings and engage in continuous professional development?			
To what extent does this teacher collaborate with others when planning lessons?			
To what extent is this teacher involved in running the school?			
To what extent does this teacher's relationship with parents help students and your school			
Commitment to Teaching			
To what extent does this teacher think that there is no finer job than teaching?			

The Extended or Restricted Professional

Is this a teacher who makes use of student or teacher centered instructional activities	Student Centered	Teacher Centered

Applying IoT and Video Image Analysis for Behavioral Studies in Endangered Species: the European Mink *(Mustela Lutreola)* as an Example.

S. Villaverde-Morcillo[1], V. Iborra[2], R. de la Fuente[2], R. García-Sánchez[1] and A. Fernández[1,2]

1. Research Center, Foundation for Ethology and Biodiversity Research (FIEB), Casarrubios del Monte, Toledo, Spain, info@fiebfoundation.org.
2. Fractalia Software, Madrid, Spain, alejandro.fernandez@fractaliasoftware.com.

Introduction

Internet of Things (IoT) is revolutionizing transversely he way humans relate to the environment. Technology is considered an important tool for easing and expedites many aspects of the daily life and is having a questionless impact in science.

SMARTfieb is a Project focused in the application of technology in studying behavior in endangered species for their conservation. We are developing a sensor platform based on the Internet of Things (IoT) and mimetized in the environment to minimized the impact in the results and avoid the observer bias.

The application of technology in wildlife behavior research has enormous advantages and applications such as the increase of quantity and quality of the data obtained the increase of the variety of data, increase of the analysys capacity and decrease of the stress produced to the animals.

In this project we have combine:

- IoT management platform, Thinking Things (Telefónica).
- Sensors network.
- Cloud Control Centers.
- Increase reality and Real Time and Historic Data web visualization.

Project Development

In the SMARTfieb Project we have deployed:

- For environment monitorization:

Type of device	Model of the device	Variable measured	Number of devices
Digital Thermometer	DS18B20	Temperature	9
Temperature-Humidity sensor	DHT22	Temperature + humidity	3
Ligth detector	LDS	Lux meter	3
Gas detector	MQ-135	NH_3	4
Water flowmeter counter	1.5" Water Flow Flowmeter Counter Hall Sensor (1-120L/min)	Volume/min	3
Barometric pressure/ altitude/ temperature	BMP180	Atmosferic pressure/altitude/temperature	3

All these devices will record environmental data to analyze their influence in behavior variation. Changes in activity periods, weight and expression of specific behavior such as nest construction, food storing or territorial marks.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

- European mink facilities.

Type of device	Model of the device	Variable measured	Number of devices
HD video camera		Video recording	20
Mobotix camera	Hemispheric Q25	Thermal track video tracking	1
RFID detector		Individual identification	20
Gas detector	MQ-135	NH3	4
Water flowmeter counter	1.5" Water Flow Flowmeter Counter Hall Sensor (1-120L/min)	Volume/min	3
Barometric pressure/ altitude/ temperature	BMP180	Atmospheric pressure/altitude/temperature	3

All de sensors and video cameras installed in the European mink facilities allow us to identify and describe the behaviors to stablish the physiological limits.

- European mink intelligent nest-box.

Type of device	Model of the device	Variable measured	Number of devices
HD video camera		Video recording	10
RFID detector		Individual identification	10
Gas detector	MQ-135	NH3	10
Temperature/humidity	DHT22	Temperature + humidity	3
Scale		Weight	10

This intelligent nest-box allow us to have a lot of information of the secret life of the European mink with a minimal disturbance. We can also weight the animal every time it pass through the first chamber of the next having an average value of several measurements which result in an accurate value, this value is very interesting for detecting weight loss due to illness or weight increase related to pregnancy.

- Control devices.
  - Videowall for video vigilance and analysis.
  - Dashboard for real time data visualization accessible on-line.

APPLYING OF THE SMARTfieb PLATFORM IN THE EUROPEAN MINK CONSERVATION.

These platform is being used for identifying normal behavior (ethogram) and stablishing the normal parameters frequency, seasonality and intensity in European mink behavior. The establishment of the normal parameters will permit us to program alerts to be aware of:

- Strange weight variations.
- Hyper- or hypoactivity caused by different conditions either physiological or pathological.
- To advance physiological events such as copula or births.
- Animal Welfare evaluation

The control and data obtained from the animals are maximum with a minimum of stress and bias.

References

1. Atzori, L., Iera, A., Morabito, G., 2010. The Internet of Things: A survey. Comput. Networks 54, 2787–2805.
2. Stevenson, M.F., Poole, T.B., 1976. An ethogram of the common marmoset (Calithrix jacchus jacchus): General behavioural repertoire. Anim. Behav. 24, 428–451.

Stratification of Human Anxiety through Biometrics and Pattern Classification

James Kah[1], Jess Low[1], Lim Pei[1], Teo Swee Teng[1], Woon Cher Xu[1], Wu Po Cheng[1], Frederick Tey[2] and Maureen Lee[2]

[1]Department of Biomedical Engineering, National University of Singapore, Singapore
[2]Combat Protection & Performance Programme, DSO National Laboratories, Singapore

Motivation

Anxiety is often a subjective and emotional state that is hard to quantify. Therefore, finding a universal criterion for quantification of stress is challenging. A method to stratify stress into manageable and indicative levels could help individuals in taking charge of their stress through various forms of interventions such as mindfulness program.

Methods

Based on literature review, Skin Temperature (ST), Galvanic Skin Responses (GSR) and Heart Rate Variability (HRV) are three of the most sensitive parameters that will vary under anxiety. After preliminary trials on human subjects and observation of the data obtained, our team hypothesized that these biometrics can be stratified into 4 distinct clusters which represent 4 different levels of anxiety. A pattern classification algorithm known as the k-Nearest-Neighbour (kNN) Algorithm was used for the stratification.

To prove our hypothesis, our team conducted a stress-inducing experiment on 25 screened male participants. The participants were subjected to 4 levels of stress; at rest, level 1, 2, 3 using The Multitasking Framework (Purple Research Solutions, UK). Their biometrics were collected and classified into distinct clusters based on the different stress levels. Each participant has a unique set of clusters. These clusters were filtered by removing those data points that falls (i) below the first quartile subtract 1.5 times of the interquartile range or (ii) above the third quartile plus 1.5 times of the interquartile range. To determine if each cluster is significantly different from the other clusters, t-test was performed and statistical significance was based on a p-value threshold of 0.05. With the use of the kNN Algorithm, human anxiety was classified into one of these clusters in which most of its neighbours lie in. NASA Task Load Index and Visual Analog were also utilized to provide subjective feedback on the progressing difficulty of each level.

Results

18 participants showed clear clusters of data points stratified according to the difficulty level of the tasks. See Figure 1 for the stratification for 1 of the 18 participants with 4 distinct clusters. 7 of the participants did not show the stratifications or did so without adherence to the difficulty. Possible reasons for diversion include focal hyperhydrosis (excessive sweating) and different perception of difficulty. For the participants who were able to show clear stratifications, the clusters were specific to the individuals and the distinction between each level was well-defined. This attest to the assumption that stress response is personal, but biometrics such as ST, GSR and HRV do respond to growing anxiety in ways specified by previous researches. Specifically, we have obtained clear stratification from GSR and ST.

There was also successful separation of the clusters into different levels (Rest, Level 1, 2 & 3) for all the participants, excluding those who have encountered experimental error. This demonstrated the validity of our stratification method via the kNN and our hypothesis to stratify human stress with 4 levels.

Based on the subjective responses obtained from the NASA Task and visual analog, subjects felt increasing workload and stress with increasing difficulty levels from the Multitasking Framework. This showed that the Multitasking Framework was able to induce varying level of stress by altering its difficulty.

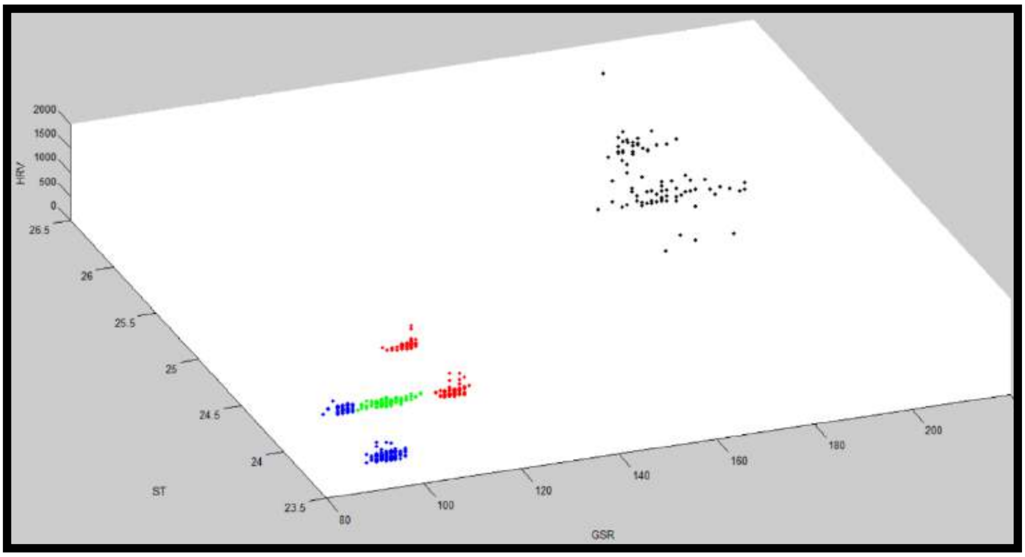


Figure 1. Stress Stratification for 1 of the Participants. Black, Red, Green, Blue Represent At Rest, Level 1 Stress, Level 2 Stress and Level 3 Stress Respectively.

Conclusion

From the findings presented, we were able to obtain promising data to bring a valid quantification method to human anxiety. With this method, individuals will be able to monitor their own stress with their own biometric responses, allowing for meaningful interventions. Further work includes investigation on raising the successful application rate of this method.

Reference

1. Purple Research Solutions (2005). The Purple Framework. < http://www.purple-research.co.uk/framework.html>. Accessed 1st October 2015.
2. Mark A. Wetherell, Kirsty Carter. (2013). *The Multitasking Framework: The Effects of Increasing.* Newcastle: Health in Action: Stress Research Group,Northumbria University.

Using Gait Analysis and Weight Bearing in Non-clinical Analgesic Research

H. Svärd[1], A. Suominen[1], J. Immonen[1], K. Ängeby Möller[2], C. Stenfors[1], J. Sallinen[1] and J. Holappa[1]

[1]CNS Research, Orion Corporation Orion Pharma, Turku, Finland. Heta.Svard@orionpharma.com
[2]Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden.

Introduction

Rats are commonly used in pharmacological research when investigating efficacy of novel analgesics. However, novel drugs that seem to relieve pain-like behaviour in rats have often not been efficacious in humans [1, 2, 3]. Assessing pain in rats is challenging, considering their prey nature of hiding pain. In contrast to measuring the magnitude of an evoked response, for instance by applying a mechanical stimulus with increasing force on the rat's painful paw, assessment of spontaneous behaviour such as walking, might give more representative information about the rat's pain state [4]. Readouts from spontaneous behavior could better model human pain and might make the translation of efficacy of novel analgesics from rat to human more reliable. Examination of joint pain in osteoarthritis patients involves assessment of pain during walking and at rest [5]. Gait analysis is typically used for studying locomotor effects of spinal cord injuries in rodents [6]. Recently, it has also been applied in studying analgesic efficacy of drugs in monoarthritic rats during locomotion [7]. Our aim was to a) adapt the findings on the related apparatus PawPrint [7] to the commercially available CatWalk™ system and to validate the method with clinically used analgesics and b) to compare walking pain to pain at standing using the Linton's incapacitance tester.

Materials and methods

Ethics statement

All studies were approved by the Regional State Administrative Agency of Southern Finland (approval number ESAVI/7238/04.10.07/2014).

Animals

A total of 72 male Wistar rats (HsdRCCHan:WIST, Harlan, The Netherlands and United Kingdom, weighing 160-230 g at the start of behavioural testing) were used in the studies. All rats were naïve to previous treatments and tests. They were housed four per cage in Makrolon IV cages with aspen chips as bedding material and a red polycarbonate tube as cage enrichment. The rats had free access to food (SDS RM1 (E) SQC, Special Diet Services Ltd, Witham, England) and tap water throughout the studies. The temperature (22 ± 2°C), humidity (55 ± 15%), ventilation (12.5 ± 2.5 times/h) and lighting of the animal room (lights on from 5.30 a.m. to 5.30 p.m. including 30 min time for gradual changing in the light condition; day 300 ± 60 lux and night 5 ± 4 lux) were continuously monitored and controlled. The animals were acclimatized to the site for at least a week before starting the experiments and habituated to the test room for a minimum of 30 min before each experiment. All experiments were carried out during the light phase.

Induction of monoarthritis

To induce inflammatory joint pain, 50 µl Freund's Complete Adjuvant (CFA; Sigma-Aldrich, containing 1 mg heat killed and dried *Mycobacterium tuberculosis* per ml) was injected into the left tibio-tarsal joint of the rats. The injection was given under deep isoflurane anaesthesia (5% isoflurane in oxygen/air). Naïve rats were used for comparison.

Behavioral testing

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

To assess changes in gait, the automated gait analysis system, CatWalk™ XT 10 (Noldus, The Netherlands) was used. In this apparatus rats are trained to walk along a glass-based corridor with a homecage at the other exit. A light, projected into the long edge of the glass, is almost completely refleced internally except when a paw is placed on the glass. The area of contact produces an image at a given position and time point, and the intensity of the light in print correlates to the force exerted by the paw. These paw prints were videotaped from underneath and computer-analysed.

To assess pain at standing, the rats were placed standing on their hind paws on an incapacitance tester (Linton Instrumentation, The United Kingdom). This setup measured static weight bearing as force exerted by each hind limb. The measured parameters are listed and explained in Tables 1 and 2. The rats were habituated to the apparatus on two daily sessions before any test.

Table 1. Incapacitance test parameters.

Parameter	Explanation
Weight bearing (g)	Weight of a hind paw during standing, measured separately for each hind paw
Weight bearing ratio	Weight of the CFA-injected divided by weight of the non-injected hind paw during standing
Weight bearing (%)	Percent weight bearing of each hind paw in relation to both hind paws during standing

Table 2. Gait analysis parameters.

Parameter	Explanation
Weight bearing of each paw	Force of a paw against the glass floor during walking, measured separately for all four paws
Weight bearing (%)	Percent weight bearing of each paw in relation to all paws during walking
Guarding index (%)	Difference in weight bearing percentage between the CFA- and non-injected hind paw during walking
Duty cycle (%)	Proportional duration of a paw placement on the glass floor during a step cycle
Regularity index (%)	Proportional number of normal step patterns
Base of support (cm)	Average distance between the hind paws during walking

Validation with drugs

Clinically used analgesics naproxen (2.5 and 7.6 mg/kg) and pregabalin (10 and 30 mg/kg) were used to validate the model and to reverse the CFA-induced pain behaviour. Naproxen (Naproxen sodium, Sigma-Aldrich) and pregabalin ((S)-Pregabalin, Toronto Research Chemicals Inc., Canada) were dissolved into 0.9% saline and administered orally in a volume of 3 ml/kg body weight. The same volume of 0.9% saline was administered to the control groups. Vehicle or drug treatments were started one day after the injection of CFA and repeated for four days. Naproxen was administered twice a day, 4 h before behavioural tests and 1-2 h thereafter, and pregabalin once a day, 2 h before the tests. The rats were allocated to treatment groups according to the Latin square method so that the number of animals was eight in all treatment groups. The experimenter was blinded to pharmacological treatments but not to the CFA injection. The behavioural tests were performed before injection of CFA to establish baseline, 24 h after the injection of CFA before administration of drugs, and then once daily after administration of drugs for a total of four days.

Statistical analyses

To determine significant effects of drugs on response measures, data were analysed with one-way analysis of variance (ANOVA) and Dunnett's post hoc test, using GraphPad Prism version 5.02. Comparisons were made against the CFA + vehicle treatment group. A P-value below 0.05 was considered significant.

Results

CFA-induced monoarthritis caused long-lasting and uneven weight distribution between the hind paws, both during walking (CatWalk) and standing (Incapacitance tester), indicated as changes in each paw's weight bearing, and consequently in weight bearing ratio and guarding index. In the CatWalk™ system, other gait parameters including duty cycle, regularity index and base of support were also changed but the effects on those were more transient with greater inter-individual variance within a treatment group. Treatment with the non-steroidal anti-inflammatory analgesic naproxen partly and dose-dependently restored CFA-induced deficits in gait at the doses tested, while pregabalin, a drug used to treat neuropathic pain, did not.

When comparing all the measured parameters between monoarthritic and naïve animals, the effect was greatest for those measuring weight distributions between the hind paws. This held true both in movement and during standing. Likewise, these weight distribution parameters appeared to be the most sensitive ones to detect analgesic effects.

Conclusions

The CatWalk™ system and Linton's incapacitance tester serve as semi-automated, objective and reproducible assessment of monoarthritic pain in the rat during walking and standing. Above all, parameters like guarding index and weight bearing ratio that measured body weight distribution between the monoarthritc and control paw were found useful. These parameters could also be partially blocked by an anti-inflammatory analgesic, but not by an analgesic having efficacy in neuropathic pain. Other parameters like duty cycle, regularity index and base of support complemented the characteristics of gait, but they may not be as useful in the assessment of joint pain. The difference in the outcomes of the drugs suggests the CFA model of inflammatory joint pain together with the measures we used here are sensitive to relevant, and not just any, pharmacological compounds. Ängeby Möller and colleagues [7] have made similar findings with other analgesics on this model using the same gait-related parameters on another gait analysis apparatus, the PawPrint. We used the commercially available CatWalk™ system and our results were consistent with theirs. Together the results support the use of rat CFA model of monoarthritis with these parameters of gait analysis and incapacitance test in translational joint pain research. We think that the use of gait analysis and incapacitance test do not rule each other out in analgesic research since the one measures pain while walking and the other at standing. Instead, they rather support one another and we suggest using them in parallel in order to get a better overview of the studied compounds. We consider these setups and parameters to serve high translational potential and predictive validity and hence to be useful tools in the search for novel analgesics.

References

1. Hill, R. (2000) NK_1 (substance P) receptor antagonists – why are they not analgesic in humans? *Trends in Pharmacological Sciences* **21**, 244-246.
2. Wallace, M.S., Rowbotham, M., Bennett, G.J., Jensen T.R., Pladna, R., Quessy, S. (2002) A multicenter, Double-Blind, Randomized, Placebo-Controlled Crossover Evaluation of a Short Course of 4030W92 in Patients With Chronic Neuropathic Pain. *The Journal of Pain* **2**, 227-233.
3. Mogil, J.S., Davis, K.D., Derbyshire, S.W. (2010) The necessity of animal models in pain research. *Pain* **151**,12-17.
4. Mogil, J.S., Crager, S.E. (2004) What should we be measuring in behavioral studies of chronic pain in animals? *Pain* **112**,12-15.

5. Lane, N.E., Schnitzer, T.J., Birbara, C.A. Mokhtarani, M., Shelton, D.L., Smith, M.D., Brown, M.T. (2010) Tanezumab for the Treatment of Pain from Osteoarthritis of the Knee. *The New England Journal of Medicine* **363**, 1521-1531.
6. Koopmans, G.C., Deumens, R., Honig, W.M.M., Hamers, F.P.T., Steinbusch, H.W.M., Joosten, E.A.J. (2005) The Assessment of Locomotor Function in Spinal Cord Injured Rats: The Importance of Objective Analysis of Coordination. *Journal of Neurotrauma* **22**, 214-225.
7. Ängeby Möller, K., Berge, O.G., Finn, A., Stenfors, C. (2015). Using gait analysis to assess weight bearing in rats with Freund's complete adjuvant-induced monoarthritis to improve predictivity: Interfering with the cyclooxygenase and nerve growth factor pathways. *European Journal of Pharmacology* **756**, 75-84.

Web Streaming of Leaf-cutter Ant Colony Activity for Outreach and Engagement

Matthew S Sullivan, Mark Peace, Sam Illingworth, Scott Pedley

Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK.

At Manchester Metropolitan University we maintain three colonies of leaf-cutter ants (*Atta cephalotes*, *Acromyrmex octospinosus*) for teaching purposes. We have a strong interest in public engagement and we have a vision of streaming live behavioural data from foraging ants to the web where it can be accessed by schools all over the world. Our medium-term goal is to engage schools by asking them to suggest experiments such as testing for forage preferences, and these experiments are then set up by MMU staff with the ant behaviour live streamed to the classroom. This poster illustrates how we have gone about achieving these aims by finding novel solutions to monitoring using open source hardware (Raspberry Pi, Arduino) and software. By keeping the systems simple and open source we can act as a hub to receive and coordinate ant colonies in education from all over the world.

The colonies are kept in glass aquarium tanks approx 20x20x40cm, and these are housed in constant environment cabinets. There are two clear acrylic tubes approx 3cm diameter exiting from each colony chamber; one goes to a waste tank in the cabinet, while the other leads out into the ambient laboratory conditions and from each colony can make a distance of around 28m. Every three metres a T-junction piece is in place in the main travel tube that allows ants to exit the tube into a side tube, where another T-piece allows connection of two forage tanks. These are typically used for forage preference experiments such as e.g. laurel vs holly leaves.

The principle data we will collect and stream will be movements of individual ants at census points in the foraging tubes. Our ant counter will at first operate by using a number of led light gates wired to an Arduino board. Still under development is an image processing ant counter which will work alongside the web cams. Ants counted are assigned to a number of categories including moving toward or away from their colony; with or without a leaf fragment. The data will also be stored as csv files on a continuous basis which will allow examination for the first time of long-term cycles of behaviour of the colonies.

The simple, inexpensive and open-source nature of this setup allows it to be replicated at little cost to educators and researchers across the globe. We invite all to join this web stream of data.

A Simplified Scan Sampling Procedure for Evaluating Behavioral Effects of a Group Intervention

S.F Broughton and C.R. Hill-Chapman

Department of Psychology, Francis Marion University, Florence, SC, USA
sbroughton@fmarion.edu

Introduction

Observational methods of assessment are integral to behavioral research. Developing a clearly defined target behavior paired with a high quality measurement procedure is the essential first step in behavioral research. Numerous observational procedures have been developed to meet the needs of researchers studying animal and human behaviors under diverse circumstances and experimental conditions. A standard method is a procedure known to behavior analysts as behavioral event sampling or frequency counting in which each occurrence of a rigorously defined behavior is continuously monitored and tallied upon its detected occurrence. Alternately, each detected instance of a behavior can be timed from onset to termination to yield a measure of behavior duration. Behavior counts can be combined with timing to provide a rate measure for the behavior of interest. Such strategies capture every instance of the target behavior at the moment and/or for the duration it occurs. Operant conditioning researchers use rate-per-minute of behavior as an extremely sensitive indicator of an organism's moment-to-moment response to environmental contingencies and stimuli. Basic laboratory operant conditioning researchers generally use response activated behavioral recording instruments to collect response data continuously, avoiding the need for human observers.

Applied researchers collecting data under real world conditions usually are not able to employ intermittently monitored, continuously operating machinery to collect behavioral observations. Because conditions in natural environments and administrative, financial, ethical, and staff constraints create additional barriers to data collection in educational or clinical settings, researchers developed a variety of creative observational strategies to meet their needs for valid and reliable data early on (Bijou, Peterson & Ault, 1968). The need for behavioral data collection methods that are economical of time and clinical staff who have other responsibilities has led to the development of observational procedures not requiring continuous behavioral monitoring. Discontinuous methods of data collection have become a preferred data gathering methodology of applied behavioral researchers (Mudford, Taylor, & Martin, 2009) only requiring an observer's attention for a moment or short period at preprogramed instances. Methods preferred over event sampling or duration measurement in applied research have been based on temporal sampling strategies such as interval recording and time sampling. Typical behavioral sampling procedures include whole interval recording, partial interval recording, and momentary time sampling. These methods have been evaluated against continuous behavioral measurement methods in simulated/lab and applied settings to determine the extent to which they produce representative depictions of continuous measures of response frequency and duration (Meany-Daboul, Roscoe, Bourret, & Ahearn, 2007; Repp, Roberts, Slack, Repp, & Berkler, 1976). Research findings indicate that whole interval methods generally underestimate the frequency or duration of a behavior, while partial interval procedures overestimate the behavior's occurrence. Momentary time sampling has not been associated with a particular direction of error. Researchers, being mindful of these methodological characteristics, can choose procedures to provide the most rigorous test of interventions (e.g. use whole interval sampling procedures when intervening to increase a lower rate behavior) (Fiske & Delmolino, 2012). Nevertheless, interval and time sampling procedures employing numerous or complex behavioral codes, short observation intervals, or long observation sessions may produce strained attention or cognitive fatigue for observers, increasing observational errors.

The above noted methods generally are employed when discrete behaviors or behavioral state(s) have been identified for study in a focal individual. An additional procedure known as scan sampling is employed when groups of individuals constitute the subjects of interest. In this procedure a social group of individuals is sampled

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

at predetermined instances in time. The procedure is much like momentary time sampling, except that instead of recording the specified behavior of an individual at the moment of the sample, an entire group is scanned and recorded at the moment of the sample. At the starting moment of the scan, the observer records the behavior of the first individual in a group and then proceeds to record the behavior of the next, then the next and the next until all individuals in the group have been observed for the scan. The observer then waits for the moment of the next scan to begin (Pellegrini, Symons, & Hoch, 2012). Individuals in the group may be identified and data preserved in individually identifiable form, or individuals may be observed randomly with data aggregated across individuals to represent only the total group for each scan. In either case, observers must be careful to observe all individuals in the identified group, and to observe each individual only once per scan. In addition, all individuals in the group must be visible to observation during each scan (Hawkes, Kaplan, Hill, & Hurtado, 1987). Scan sampling procedures with short scanning intervals, numerous individuals to scan, numerous required scans, or multiple behaviors or categories can be cognitively taxing on observers, increasing the likelihood of errors. A simplified and less challenging scanning method involves scanning the group to determine whether at least one individual is engaged in the behavior of interest at the moment of the scan (Gilby, Pokempner, & Wrangham, 2010).

Development and Use of the Scan Sampling Procedure

We developed a scan sampling procedure to evaluate a preschool group intervention using a variation of the simplified procedure described by Gilby, et al. (2010). The scan sampling procedure was needed to evaluate the effect of a preschool intervention curriculum (The Promoting Alternative Thinking Strategies [PATHS] program) (Kusché & Greenberg, 1994) for increasing appropriate group engagement in children enrolled in 3K and 4K classrooms. Previous research on this intervention program used teacher informant ratings, rather than direct behavioral assessment, to determine whether prosocial/attentive behaviors increased and disruptive/aggressive behaviors decreased in the classroom. We examined the items contained in the teacher rating forms employed by the PATHS program to develop behavioral definitions for two mutually exclusive categories of behaviors: Concentration/Attention related behaviors (e.g., sits/remains in assigned location, looks at/attends/listens to instruction, maintains attention/focus, engages in activities/follows instructions, participates in/completes tasks, etc.) and Aggression/Disruptive Behavior (e.g., leaves assigned location, engages in/attends to uninstructed activities, yells at others, fights/pushes/touches others inappropriately, refuses instructions/engages in prohibited behaviors, etc.). The Concentration/Attention category included behaviors represented by the Concentration/Attention and Social and Emotional Competence sections of the PATHS TEACHER SOCIAL COMPETENCE RATING SCALE. The Aggression/Disruptive Behavior category included behaviors represented on the Aggression/Disruptive Behavior section of the PATHS TEACHER SOCIAL COMPETENCE RATING SCALE.

Students from 3K and 4K classrooms at a low-income childcare facility participated, resulting in a sample of 23 students. Thirteen of the students were from the 3K class, and ten of the students were from the 4K class. Six of the students were male (26%), seventeen were female (74%); three of the students were Asian (13%), eleven were Caucasian (48%), five were African American (22%), and four were Bi-Racial (17%). The ages of the students ranged from three years of age to five years of age; all fell within the preschool range.

Observations were made during randomly selected morning and afternoon group/circle time structured activities during baseline (Baseline) as well as during implementation of the PATHS program (PATHS Generalization) and after the program had ceased (Follow-Up), and during PATHS lesson sessions during implementation of the intervention program (PATHS). Thirty-nine observation sessions were completed in the 4K class and 28 in the 3K class. A fifteen second interval group momentary scan sampling procedure was employed. At the beginning of each observation interval an auditory cue was sounded. At the moment of the cue, observers scanned the group and determined whether all members of the group were on-task/attending/concentrating or not. If all children were appropriately engaged during the scan, a plus (+) was circled for the interval on the observation form. If any one or more of the children in the group were not appropriately engaged or were engaged in aggressive/disruptive behavior, then a minus (-) was circled on the observation form for the interval. Thus, our procedure differed from the simplified procedure described by Gilby, et al. (2010) in that we recorded the occurrence of concentration/attention behaviors only if *all* members of the group (rather than just one) conformed to the behavioral definition. This provided a rigorous test of the intervention in that all children had to be in compliance for the group to be scored as complying with the behavioral definition during any given interval. Observations were conducted for 20 minutes per session, or until the end of the activity, whichever occurred first. At the end of each observation session, the percentage of observation intervals of on-task/concentration/attention behavior was computed by dividing the number of intervals recorded as a plus (+) by the total number of observation intervals recorded.

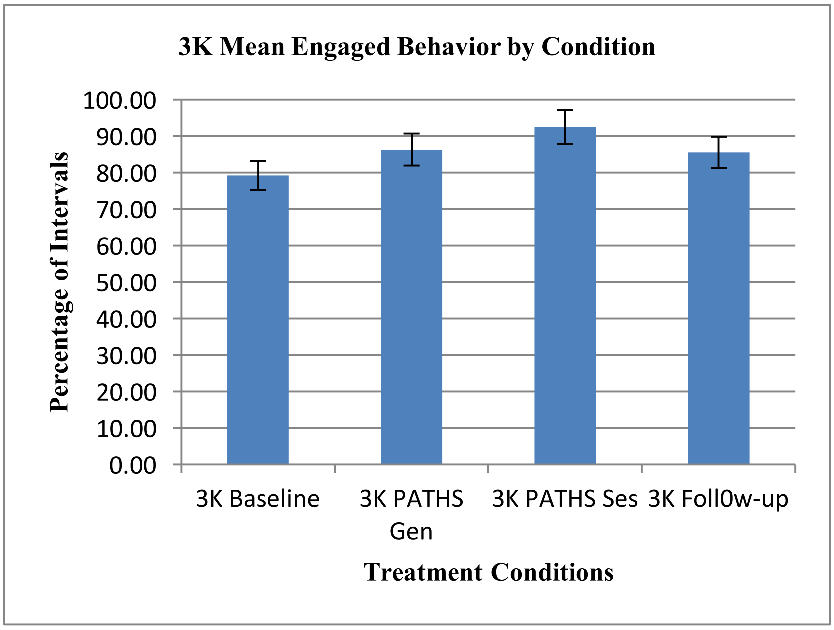
One university faculty member (supervisor) and two university students served as observers. Both students had completed coursework in behavioral assessment and analysis. Observers were provided written copies of the behavior definitions and coached to observe in pairs with the supervisor until all possible pairs obtained 80% inter-observer agreement or higher. Inter-observer agreement was computed using the exact agreement method. Reliability computations were based only on intervals in which one or both observers recorded a plus (+). Agreements were scored when both observers recorded plusses, and disagreements were scored when one observer scored a plus and one a minus. Observer agreement was computed using the formula: [Agreements/ (Agreements + Disagreements)] x 100. In the 4K class, reliability observations were conducted on 18% of the observation days spread throughout the conditions of the study. Mean inter-observer agreement for the 4K class was 86% with a range of 78 to 92%. Reliability observations were conducted during 14% of observation days in the 3K class spread throughout the conditions of the study. Mean inter-observer agreement for the 3K class was 95% with a range of 92 to 100%.

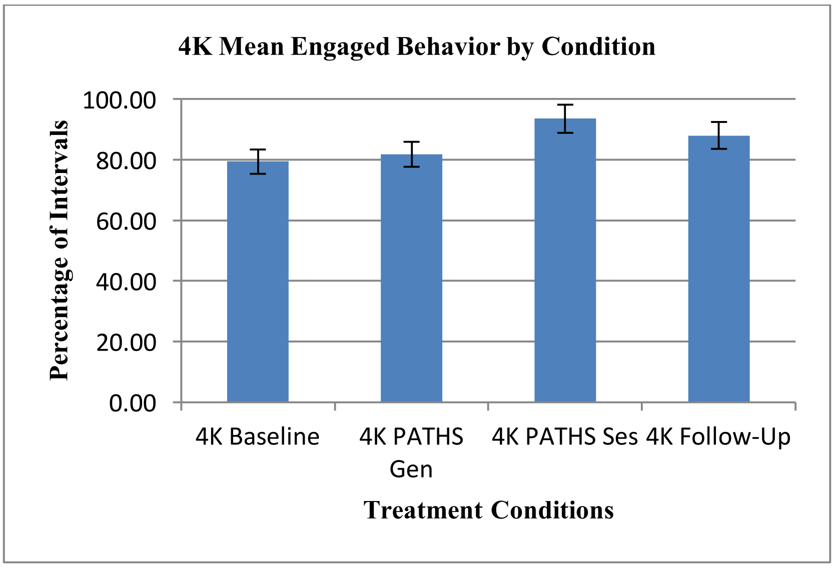
Results

Mean percentage of intervals of engaged behavior per treatment condition for each group is shown in Table 1 and Figures 1and 2. The momentary scan sampling procedure was consistently sensitive to changes in group response to each treatment condition. Classroom engagement varied systematically with treatment conditions. In order to quantify the magnitude of the change in level of performance, we used a variation of Cohen's (1988) d statistic as calculated by Busk and Serlin (1992): where the mean for pre-treatment (Baseline) is subtracted from the mean for post-treatment (Follow-up) and divided by the standard deviation for the pre-treatment period. Empirical assessment to date indicates that Busk and Serlin's d statistics is the most reliable estimator of the effect size when the pre-treatment variance is a non-zero value for within-subject repeated measures studies (Beeson & Robey, 2006). Based on these calculations, we determined the effect sizes for the 3K and 4K classrooms to be .52 and .66, respectively. These results are consistent with the pre-intervention and post-intervention teacher ratings of the students in each class in which statistically significant reductions in emotional symptoms, conduct problems, hyperactivity-impulsiveness, and peer problems were obtained along with statistically significant increases in prosocial behavior.

Class	**Value**	**Baseline**	**PATHS Generalization**	**PATHS Session**	**Follow-up**
3K					
	Mean	79.11	86.23	92.50	85.50
	Std. Dev.	12.29	12.14	6.61	10.61
	Std. Error	4.10	3.37	2.34	7.50
	95% conf.	8.03	6.60	4.58	14.70
4K					
	Mean	79.36	81.78	93.55	88.00
	Std. Dev.	13.09	9.70	5.61	0.00
	Std. Error	3.95	2.02	1.69	0.00
	95% conf.	7.74	3.97	3.32	0.00

Table 1: Mean and Variability of Percentage of Intervals by Treatment Condition





Figures 1 and 2: Mean Percentage of Intervals of Engaged Behavior by Condition and Class

Discussion

This study used a simplified momentary scan sampling procedure to obtain direct real-time measurement of children's engagement in preschool class activities. This represents an improvement in program measures, which previously had consisted only of pre- and post-treatment teacher ratings of the children. The scan sampling method allowed visualization of each group's response to each treatment condition over time in addition to providing a pre- post-assessment. The momentary scan sampling procedure proved to be reliable, and the overall program evaluation from the scan sampling data was consistent with the teacher rating data also collected. Results were consistent with the program's hypothesis that implementation of the program would increase classroom engagement. Although effect sizes were relatively small, the results are promising. To the researchers' knowledge, this is the first study that has attempted to measure the effect of PATHS on behavioral measurement of classroom immediate

References

1. Bijou, S.W., Peterson, R.F., & Ault, M.H. (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis,**1**(2),* 175-191.
2. Mudford, O.C., Taylor, S.A., & Martin, N.T. (2009). Continuous recording and interobserver agreement algorithms reported in the *Journal of Applied Behavior Analysis (1995-2005). Journal of Applied Behavior Analysis, **42***, 165-169.
3. Meany-Daboul, M.G., Roscoe, E.M., Bourret, J.C, & Ahearn, W.H. (2007). A comparison of momentary time sampling and partial-interval recording for evaluating functional relations. *Journal of Applied Behavior Analysis*, ***40***(3), 501-514.
4. Repp, A.C., Roberts, D.M., Slack, D.J., Repp, C.F., & Berkler, M.S. (1976). A comparison of frequency, interval, and time-sampling methods of data collection. *Journal of Applied Behavior Analysis*, ***9***(4), 501-508.
5. Fiske, K. & Delmolino, L. (2012). Use of discontinuous methods of data collection in behavioral intervention: Guidelines for practitioners. *Behavior Analysis in Practice*, ***5***(2), 77-81.
6. Pellegrini, A.D., Symons, F., & Hoch, J. (2013). *Observing children in their natural worlds: A methodological primer*. New York: Psychology Press.
7. Hawkes, K., Kaplan, H., Hill, K., & Hurtado, M. (1987). A problem of bias in scan sampling. *Journal of Anthropological Research*, ***43***(3), 239-246.
8. Gilby, I.C., Pokempner, A.A., & Wrangham, R.W. (2010). A direct comparison of scan and focal sampling methods for measuring wild chimpanzee feeding behaviour. *Folia Primatol (Basel),* ***81***(5), 254-264.
9. Kusché, C.A. & Greenberg, M.T. (1994). *The PATHS Curriculum*. Seattle: Developmental Research and Programs.
10. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
11. Busk, P.L. & Serlin, R. (1992). Meta-analysis for single case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187-212). Hillsdale, NJ: Lawrence Erlbaum Associates.
12. Beeson, P. M. & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16,* 161-169. doi:10.1007/s11065-006-9013-7.

Automated and High-Throughput Zebrafish Embryonic Behavioral Screening Platforms

Daniel Marcato[1], Eduard Gursky[1], Johannes Stegmaier[2], Ruediger Alshut[4], Maryam Shahid[1], Helmut Breitwieser[2], Robert Geisler[1], Christian Pylatiuk[2], Jos van Wezel[3], Ralf Mikut[2], Uwe Straehle[1], Ravindra Peravali[1]

[1]Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Germany
[2]Institute of Applied Computer Science, Karlsruhe Institute of Technology, Germany
[3]Steinbuch Center for Computing, Karlsruhe Institute of Technology, Germany
[4]MCC - Measure, Check & Control Gmbh & Co. KG, Karlsruhe, Germany

In recent years zebrafish (*Danio rerio*) embryonic behavior has emerged as a robust tool for investigating toxicological effects of compounds and cognitive and motor abilities and disorders. Owing to its transparency, fecundity and genetic capabilities, including transgenesis and gene knock-out, the zebrafish is amenable to phenotypic and to forward and reverse genetic screening. However, the zebrafish model becomes relevant only if a large number can be screened for phenotypes in a fast and unsupervised manner. Here we present three behavioral screening platforms that are completely automated starting from sample handling, imaging and finally data analysis and storage.

One of the bottlenecks in using zebrafish embryos for high-throughput studies is the manual handling and plating of the embryos. We present a "zebrafish sorting robot" that is capable of pipetting zebrafish eggs and larvae (up to 120 hours post fertilization (hpf)) into 96- and 384- microtiter plates and into petri-dishes. The zebrafish sorter is the first component of all our behavioral screening platforms.

The first behavioral imaging system is the high-throughput Photomotor Response (PMR) system that aims to characterize one of the earliest manifestations of behavior in zebrafish: the response of zebrafish embryos at 30 – 40 hpf to flashes of bright light [1]. Specifically, the PMR is used to study the effects of neuroactive substances and quantifies the motion of embryos within the chorion prior to, during and after presenting a series of light stimuli. This PMR platform is unique as it can be completely remotely controlled and thereby ensuring no human interference and extraneous sound and vibrations [2]. The second system, called the robotic imaging system, is equipped with a robotic arm and a high-resolution camera and is used to image morphological developmental phenotypes as well as long term locomotor and feeding behavior. This system is capable of unsupervised data acquisition from 0 – 5 days post-fertilization (dpf) of zebrafish embryos. The third system, the multi-camera parallel microscope, is equipped with multiple high-resolution cameras that are mounted on a robot that moves in the X-Y plane. The presence of multiple cameras as opposed to a single detector significantly reduces the time to scan a multi-well plate and can also simultaneously scan multiple petri-dishes. This system can be robustly used for simultaneous long-term time-lapse imaging of zebrafish embryos.
We will present a pilot study using some known neuroactive compounds and will discuss the performance of our screening platforms. In addition, we will present our specialized real-time and quantitative image analysis and processing pipelines. These high-performance computing pipelines extract locomotor kinematics and behavioral dynamics. Finally, since behavioral experiments are data intensive we will present the data intensive computing infrastructure at the Karlsruhe Institute of Technology (KIT) that includes the Large Scale Data Facility (LSDF) with several peta-bytes of storage and archival capacity.

Declaration

All experiments used zebrafish larvae between 1 and 5 dpf for which no special animal licensing is required and was carried out following European and national regulations. Husbandry is performed in accordance with the German law on Animal Protection and the Directive 2010/63/EU. Our animal facility is supervised and regularly inspected by the competent authority "Regierungspräsidium Karlsruhe" (#Az.35-9185.64).

References

1. Kokel, D. and Peterson ,R. T. (2011). Using the zebrafish photomotor response for psychotropic drug screening. *Methods Cell Biol.* **105**, 517-24.
2. Marcato, D., Alshut, R., Breitweiser, H., Mikut, R., Straehle, U., Pylatiuk, C. and Peravali, R. (2015). An automated and high-throughput photomotor response platform for chemical screens. *Proc. IEEE Eng. Med. Biol. Soc.*, 7728 – 7731.

Ethical Statement: All experiments used zebrafish larvae between 1 and 5 days post fertilization for which no special animal licensing is required and were carried out following European and national regulations. Husbandry is performed in accordance with the German law on Animal Protection and Directive 2010/63/EU. Our animal facility is supervised and regularly inspected by the competent authority "Regierungspräsidium Karlsruhe" (# Az.35-9185.64).

RumiWatch – An Automatic Sensor-based Measurement of Chewing Behavior in Cows and Horses

Jessica Anna Werner[1,2,3], Christina Umstatter[2], Nils Zehner[2], Matthias Schick[2]

[1]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland
[2]Agroscope, Institute for Sustainability Sciences, 8356 Ettenhausen, Switzerland
[3]University of Hohenheim, Institute for Agricultural Engineering, 70599 Stuttgart, Germany

Abstract

In modern livestock farming and animal husbandry, feeding conditions contribute highly to animal health and welfare. Therefore, chewing and activity behaviour are suitable parameters for monitoring health, nutritional and behavioural conditions in farm animals. Most previous studies concerning feed intake and chewing activity were based on visual observations only. With the development of an automated measurement system it is possible to record quantitative data. The RumiWatchSystem (ITIN+HOCH GmbH, Switzerland) comprises a noseband sensor, pedometer and evaluation software (Fig. 1).

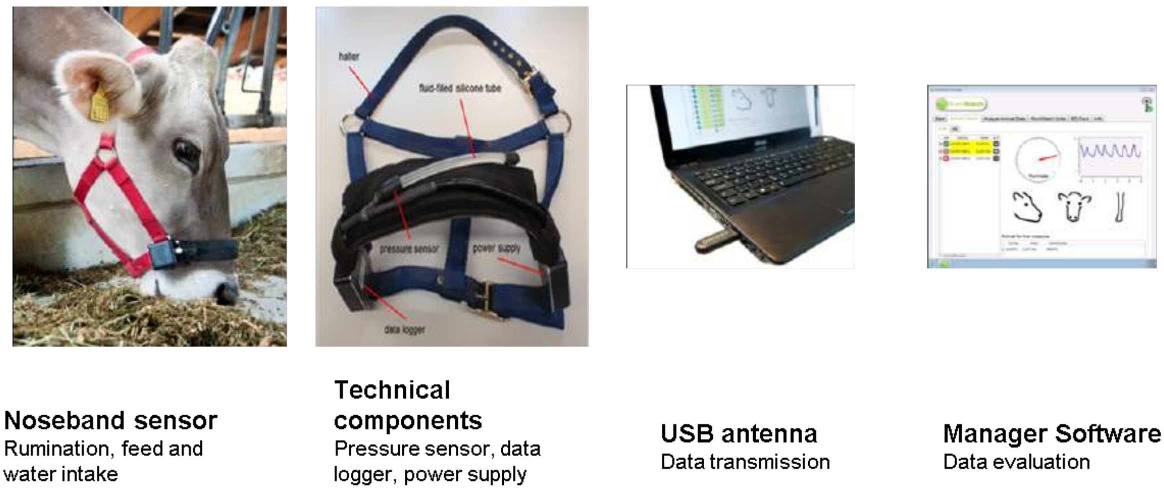


Fig. 1. The components of the RumiWatchSystem.

The RumiWatch halter consists of an oil-filled silicon tube with pressure sensor, which is integrated in the noseband of a conventional halter, as described by Nydegger at al. [1]. The signals are recorded due to a pressure change in the tube and will be saved to the data logger, which is incorporated into a plastic box at the end of the nosepiece on the right hand side (Fig. 1). The logging rate of the sensor is 10 Hz and raw data are saved to a SD memory card located in the plastic box on the right hand side. The second box (on the left) is used as a protective cover for the power supply, a 3.6 V battery, lasting up to 3 years under laboratory conditions due to a low energy operating system. Raw data can be downloaded via a USB plug-in connection. The second option of data access can be achieved by a remote download via ANT-Antenna to the software package. The raw data get classified as rumination, eating and drinking as well as other activity by using generic algorithms. All these parts enable a simple use of the halter under experimental or practical conditions.

The RumiWatchSystem was developed for indoor housing cows and is already used as a research tool [2]. In a validation study of Ruuska et al. [4], with the firmware version V01.13 and the data converter version 0.7.0.0, the noseband sensor was compared with constant video observation in rumination and eating behavior. 6 cows in total, one primiparous and five multiparous, were monitored by two observers on a 1s resolution over periods of 12 hours per cow. The results of comparing video observations with the automatic measurement system were

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

highly encouraging with R^2 for eating 0.94 and R^2 for rumination 0.93. Besides the usage in indoor housing systems, the RumiWatchSystem is also used in pasture-based systems [3].

Additonally to the use of the RumiWatchSystem in cows, there is also great potential to apply it to horses (EquiWatch). In a first study, ten horses were monitored while feeding two different types of feed (hay and haylage). Visual observations for counting chews were undertaken. The horses were watched ten minutes twice a day over a period of three days. The results were promising with a 95% accordance, although the system was not adapted specifically to the species horse [5].

Fig. 2. Application examples for the RumiWatchSystem.

Some important research applications for the RumiWatchSystem in the future could be:

- Measuring device to detect health and welfare problems.
- Measuring chewing behavior as welfare indicator.
- Measuring bites and chews on a grass based diet as a parameter for grass availability.

References

1. Nydegger, F., Gygax, L., & Wendelin, E. (2011). Automatic measurement of jaw movements in ruminants by means of a pressure sensor. *Agrarforschung*, **2**. 60-65.
2. Pahl, C., Haeussermann, A., Grothmann, A., Mahlkow-Nerge, K. & Hartung, E. (2015). 8.4. Ability to estimate feed intake from presence at feeding trough and chewing activity. *Precision livestock farming applications, edited by: Ilan Halachmi*, 311-318.
3. Rombach M., Münger A., Südekum K-H., Schori F. (2015). Validation of a new monitoring system (RumiWatch®) for recording the grazing and rumination behaviour of dairy cows *In: Proceedings of the second dairycare conference 2015*. 3-4 March, Publ. DairyCare cost action, Cordoba. 2015, 31.
4. Ruuska, S., Kajava, S., Mughal, M., Zehner, N. & Mononen, J. (2016). Validation of a pressure sensor-basedsystem for measuring eating, rumination and drinking behaviour of dairy cattle. *Applied Animal Behaviour Science*, 174, 19-23.
5. Werner, J., Umstatter, C., Zehner, N., Niederhauser, J. J. & Schick, M. (in press). Validation of a sensor-based automatic measurement system for monitoring chewing activity in horses. *Livestock Science*.

Longitudinal Characterisation of Gait Parameters in the Murine P301S tauopathy Model of Alzheimer's Disease

H. Van Craenendonck[1], S. Embrechts[1], L. Ver Donck[1], S. Beunen[2]

[1]Neuroscience Drug Discovery, Janssen Research & Development, a division of Janssen Pharmaceutica NV, 2340 Beerse

[2]Faculty of Psychology and Educational Sciences, KULeuven, 3000 Leuven, Belgium.

Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterised by cognitive decline in patients and the presence of pathological intracellular neurofibrillary tangles and extracellular amyloid plaques.

P301S MAPT (Microtubule-Associated Protein Tau) transgenic mice are a widely used model to investigate the role of aberrant tau protein in Alzheimer's disease. The *h*Tau.P301S model is characterised by hyperphosphorylation of tau and the development of age-dependent conformational changes leading to neurofibrillary tangle-like pathology in the hippocampus, cerebral cortex, brain stem, and spinal cord [1]. At an age of around 4 months these mice develop motor function deficits including clasping of the hind legs and impaired performance on the rotarod [2].

The aim of this study was to investigate whether gait parameters can be used for longitudinal characterisation of the motor phenotype of the *h*Tau.P301S mouse model.

Materials and methods

Subjects

We obtained hetero- (HE) and homozygous (HO) P301S mice (Thy1-*h*Tau: CBA.C57BL/6) from internal breeding. They were group housed per litter under standard housing conditions with 12/12 h light/dark cycle. Food and water were available *ad libitum*. The protocol was approved by the Institutional Ethical Committee on Animal Experimentation, according to applicable regional law.

Apparatus

The CatWalk 7.1 system (Noldus Information Technology, Wageningen, The Netherlands) was used for gait analysis. A horizontally mounted glass plate, illuminated at the long edge by green fluorescent light, forms the base of a 1-meter long walkway. This walkway consists of 2 parallel black Perspex plates with an interspacing of 8 cm. The home cage was placed at the end of the walkway to stimulate the mouse to cross it. The reflection of the green light illuminates the animal's plantar surfaces in contact with the glass plate, producing a bright footprint image. Video images were recorded from below while the subject traversed the glass plate. The CatWalk software was used to detect these footprints, manually label them and to calculate gait parameters.

The following static and dynamic gait parameters were calculated:

- Diagonal support: when support is on a diagonal paw pair.
- Base of support (BOS): the average width between either the front or the hind paws.
- Regularity Index: expresses the number of normal step sequence patterns relative to the total number of paw placements. This is a fractional measure of inter-paw coordination. In healthy, fully coordinated animals its value is 100%.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

- Phase Dispersions (girdle): describes the temporal relationship between the placements of two paws within a step cycle. It is used as a measure of inter-paw coordination. In normal animals the diagonal paw pairs move in synchrony, resulting in a value of 0%. Paw placement alternates in the girdle pairs (front or hind, left and right paws) resulting in a 50% value.

- Print width: width of the foot print.

- Duty cycle: As stance (duration of contact of the plantar surface with the glass plate) depends on the animal's walking speed, the duty cycle expresses stance duration as a percentage of the duration of the step cycle. A higher duty cycle means that the paw is in contact with the glass plate for a longer time, relative to the walking speed.

Procedure

The mice were repetitively tested at the age of 3, 4, 5 and 5.5 months. During experimentation, the mice were placed individually in cages in the testing room with dimmed light conditions (red light only) 30 minutes before the training and the test session.

Animals were trained to walk on the walkway prior to the test session at 3 and 4 months of age. Training comprised of consecutive trials where the subjects were gently guided along the walkway to their home cage. This was repeated until the mice independently completed five spontaneous runs. From age 5 months onwards, training was no longer needed. After training the mice were placed back in their home cage and returned to the holding room with a normal day/night light cycle for 2 hours prior to testing.

During the test phases the mice had to complete three valid runs. A valid run was defined as one in which the mice ran uninterrupted to their home cage without hesitating, stopping or returning. After the test the mice were returned to their home cage.

Analysis

Gait parameters calculated by the CatWalk software were analysed for genotype (heterozygous versus homozygous) and age effects by repeated measures ANOVA using InVivoStat statistical software (version 3.1.0.0, Bate and Clark (2014) [3]). P-values <0.05 were considered to represent statistically significant differences.

Results

As an example of changes observed in certain gait parameters of the P301S-mice, figures 1 and 2 show the results for base of support and diagonal support respectively in male HE and HO P301S-mice. For the parameter 'base of support' in the hind paws a main effect of age ($F_{(3,39)}$=8.24, p=0.0002) and a genotype*age interaction ($F_{(3,39)}$=7.12, p=0.0006) were observed. Post hoc analysis (Figure 1) showed that 'base of support' was decreased in HO-mice at the age of 5 (p<0.05) and 5.5 (p<0.001) months, while no change of this parameter was observed in HE-mice. For 'diagonal support' main effects of genotype ($F_{(1,14)}$=11.51, p=0.0044) and age ($F_{(3,39)}$=15.76, p<0.0001) were detected, as well as a genotype*age interaction ($F_{(3,39)}$=6.36, p=0.0013). Post hoc analysis (Figure 2) showed that 'diagonal support' in HO-mice was lower than in HE-mice at all ages tested (p<0.01). Furthermore, there was an age-dependent decrease in HO-mice reaching statistical significance from 5 months old onwards, while no change of this parameter was observed in HE-mice.

A summary of the statistical analysis of the other parameters is summarised in Table 1. A genotype effect was also detected for 'base of support' of front paws ($F_{(1,14)}$=11.92, p=0.0039) and for the 'regularity index' ($F_{(1,14)}$=5.41, p=0.0355). In addition, interaction effects of genotype*age were detected for 'width of the foot print' ($F_{(3,39)}$=8.05, p=0.0003) and 'duty cycle' ($F_{(3,39)}$=11.19, p<0.0001) of the front paws. On the other hand, 'phase dispersions girdle' remained unchanged.

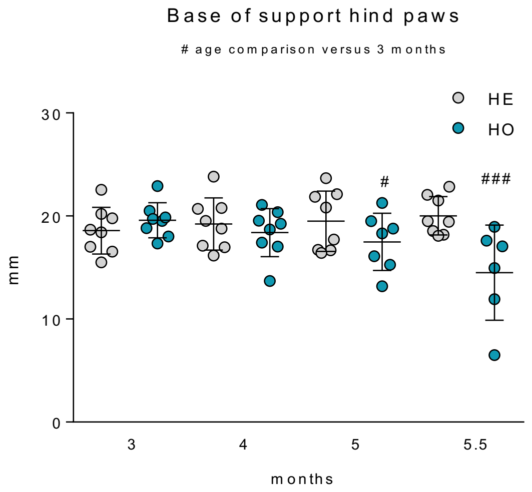


Figure 1: Base of support of the hind paws in male P301S-mice (n=6-8) tested at the different ages. Post hoc analysis shows a significant difference of age in heterozygous (HE) mice starting from five months onwards (#: p<0.05; ###: p<0.001; HE=heterozygous, HO=homozygous).

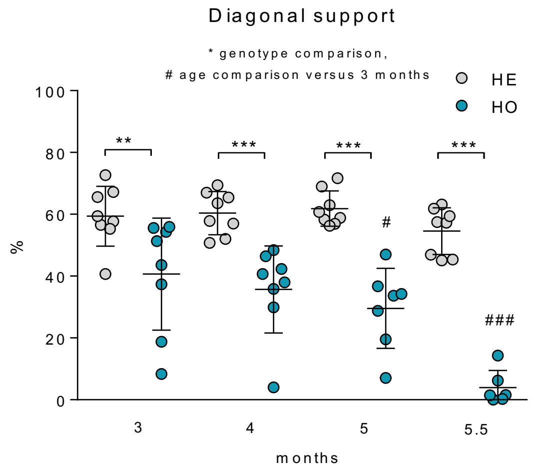


Figure 2: Diagonal support in male P301S-mice (n=6-8) tested at the different ages. Results show a significant difference between genotypes from the earliest time point (age 3 months) onwards (**: p<0.01; ***: p<0.001). There is also a significant decrease in the homozygous animals as they become older (#: p<0.05; ###: p<0.001; HE=heterozygous, HO=homozygous).

Table 1: Results of Repeated Measures ANOVA for selected gait parameters (p-values; bold indicates statistically significant effect

Parameter	Genotype	Age	Genotype × age
Diagonal support (%)	***0.0044***	***< 0.0001***	***0.0013***
Base of support, front paws (mm)	***0.0039***	0.6637	0.5585
Base of support, hind paws (mm)	0.4653	***0.0002***	***0.0006***
Regularity Index (%)	***0.0355***	0.2260	0.3861
Print width, front paws (mm)	0.5158	0.5860	***0.0003***
Duty cycle, front paws	0.5196	0.1827	***< 0.0001***
Phase dispersion girdle, front paws	0.3956	0.8004	0.1726
Phase dispersion girdle, hind paws	0.4002	0.9058	0.4398

Conclusion

Previous studies have reported impairments in motor function of P301S homozygous mice [2]. The findings in this study for the first time report changes in gait parameters in these mice, mainly related to the base of support and the degree of diagonal support. The present study did not include wild type mice as controls since these were not available as littermates from the breeding. It is important to note that HE-mice showed no age-dependent changes in the parameters analysed (data not shown). This suggests that HE-mice can be used as surrogate controls vs. the HO-mice.

Some of these parameters show changes in HO-mice vs HE starting around 5 months of age. This is in the same age range as reported for changes in motor function on the rotarod, or the occurrence of hind leg clasping [2], and suggests that these changes are correlated. Of particular interest, however, is that e.g. diagonal support was already decreased in HO-mice vs HE-mice at the earliest time point measured, 3 months of age, and further declined with age. In this perspective gait analysis can be considered as a valuable method to detect early stages of tauopathy in P301S mice.

Next steps for the longitudinal characterisation of gait parameters in P301S-mice can include testing at earlier time points, because some parameters there was already significantly different between the HE- and HO-animals. A study including wild type mice in the comparison with HE- and HO-mice should further characterise whether signs of developing tauopathy can be detected in HE-mice as well.

References

1. Takeuchi H, Iba M, Inoue H, Higuchi M, Takao K, Tsukita K, Karatsu Y, Iwamoto Y, Miyakawa T, Suhara T, Trojanowski JQ, Lee VMY, Takahashi R.. P301S Mutant Human Tau Transgenic Mice Manifest Early Symptoms of Human Tauopathies with Dementia and Altered Sensorimotor Gating. PLoS ONE 6:e21050, 2011.
2. Scattoni ML, Gasparini L, Alleva E, Goedert M, Calamandrei G, Spillantini MG. Early behavioural markers of disease in P301S tau transgenic mice. *Behav Brain Res.* 208:250-257, *2010*.
3. Bate ST, Clark RA. The Design and Statistical Analysis of Animal Experiments. Cambridge University Press, Cambridge, 2014.

Standardized Ligth/dark Preference Test for Anxiety and Stress Research Using Zebrafish Larvae

Ainhoa Alzualde and Aintzane Alday

BBD BIOPHENIX-BIOBIDE, Spain

Zebrafish has emerged as an exceptional animal model for many research and applications in several fields. The advantages of the zebrafish are mainly their low cost and ease of maintaining and breeding. Moreover, many of the research conducted in zebrafish has been carried out in larvae to take advantage of zebrafish fecundity, larvae small size, ease of handling and transparency. In addition, the use of zebrafish embryo/larvae is in accordance with the 3R principle. There are more and more behavioral assays traditionally used in rodents that are being adapted to zebrafish, firstly in zebrafish adults and afterwards in embryos/larvae.

There is an increasing number of papers supporting the usefulness of zebrafish model for anxiety and stress research. Light/Dark preference Test is one of the assays that has being adapted for behavior measurement in zebrafish [1][2]. Although rodents prefer the dark compartment to the lighted one in the laboratory, in case of adult zebrafish there are controversial results that are supposed to be conflicting because of the differences in the testing equipment [3].

In case of zebrafish larvae there is only one paper which aimed to adapt the light/dark preference test for young juvenile (6dpf - days post fertilization) [4]. In this work they used a homemade apparatus that allows testing one individual each time. Therefore, the main objective of the present project is to characterize 4-5dpf larval behavior in the light/dark preference test in a standardized system. Embryos are placed in 8-square well plates (Nunc^{TM}, USA) that is combined with a specific mask designed by Noldus to shadow the half of each well. The plate is introduced in the Daniovision system powered by Ethovision XT9 to assess the total distance moved as well as the duration in light and dark areas of each well.

In this work we will present the preliminary data of a standardized light/dark preference test for zebrafish larvae which will be easily repeatable at different laboratories in order to avoid controversial results obtained using differently designed homemade testing apparatus.

Reference

1. Serra, E.L., Medalha, C., Mattioli, R. (1999) Natural preference of zebrafish (*Danio rerio*) for a environment. *Braz. J. Med. Biol. Res*. 32(12):1555-1553.
2. Gerlai, R. (2003) Zebra fish: an uncharted behavior genetic model. *Behav. Genet*. 33(5): 461-468.
3. Blaser, R.E., Peñalosa, Y.M. (2011) Stimuli affecting zebrafish (*Danio rerio*) behavior in the light/dark preference test. *Physiology & Behavior* 104: 831-837.
4. Steenberger, P.J., Richardso,n M.K., Champagne, D.L. (2011) Patterns of avoidance behaviours in the light/dark preference test in young juvenile zebrafish: A pharmacological study. *Behavioural Brain Research* 222:15-25.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Automated Face Recognition in Rhesus Macaques

C. L. Witham[1,2]

[1] Institute of Neuroscience, Newcastle University, Newcastle-upon-Tyne, UK. c.l.witham@ncl.ac.uk
[2] Centre for Macaques, Medical Research Council, Salisbury, UK

Background

The use of automated behaviour systems for monitoring animal health and welfare is rapidly expanding field of study for a range of species including farm, laboratory and zoo animals. A major challenge in measuring the behaviour of individuals housed in groups is to reliably identify each individual. One solution is to add a tracking device to each animal and another is to use biometric identification based on the distinguishing visual characteristics of that species (e.g. coat pattern)[1]. Here I present face recognition as a means of non-invasively identifying individual rhesus macaques in videos. Rhesus macaques are one of the most common primates species used in biomedical research and there is a need for better methods of monitoring their health and welfare. Rhesus macaques do not have obvious individually identifiable features but they are capable of recognising conspecifics by their faces[2]. This raised the possibility of using face recognition to identify individuals, which is widely used in humans but its use in animals has so far been limited. However a study in apes has shown that it is possible to use automated face recognition to identify individual chimpanzees[3]. Although face recognition cannot be used to directly measure behaviour we can use face recognition in a number of ways to assist with measuring behaviour. Firstly we can use it to save time during manual coding of behaviour by identifying video clips containing a particular animal. Secondly we can use it to investigate social associations within a group by identifying animals that spend time in close proximity (discussed at the end of this paper). Finally in the future it may be possible to integrate face recognition with automated behaviour analysis to produce a fully automated system.

Methods

All methods presented were developed using group-housed rhesus macaques (*Macaca mulatta*) at a breeding facility. The monkeys were housed in groups of 6-25, in large indoor enclosures with high levels of enrichment and access to natural light (ages in the range 0-20 years). The housing exceed the national guidelines and all necessary approvals (including ethical) were given. Video footage of four different groups was collected using a high-definition video camera (Sony HDR-SR12E). The videos were then converted to MP4 format and annotated with the date and group information. All analyses was carried out in Matlab (www.mathworks.com).

Face detection

Before we can start face recognition we need to be able to detect a macaque face in an image or video frame. This is done using face detection. A cascade classifier[4] based on local binary patterns was trained on frontal facial images manually cropped from video stills. The positive training set included male and female macaque faces of all ages (examples in Figure 1A). The negative image set included images of monkeys not facing toward the camera and images of the housing enclosures. This was successful at detecting 94% of frontal faces in a test set (compared to manual analysis) but also had a high level of false positives (15% of detected faces were not actually faces) and poorly aligned faces so I implemented a second stage to the detection process to check the alignment of the eyes and nose. For this stage additional cascade classifiers were trained to find eyes and noses (example images in Figure 1A) within the confines of the detected face (Figure 1B-C). This reduced the false positive rate to less than 0.005%.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

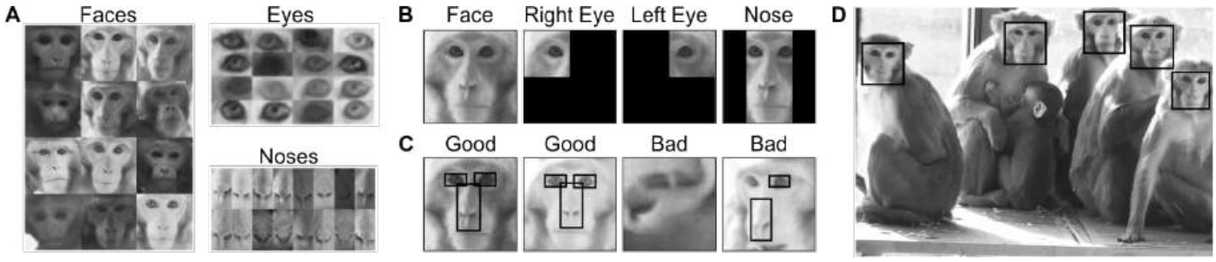


Figure 1. Face detection. A, examples of images used to train face, eye and nose cascade classifers. B, regions of face run through eye and nose cascade classifers. C, images accepted (Good) or rejected (Bad) by alignment stage. D, face detection run on a single video still, black boxes indicate detected faces.

Face recognition

The face detection described above was used to extract over 100,000 facial images from videos of four different groups of monkeys (each video was processed at a rate of 2 frames per second; single frame example shown in Figure 1D). The faces were manually sorted according to the identity of the monkey and a subset of images per monkeys were further sorted by the quality of the facial image. I created a test set for testing face recognition containing 50 well-aligned images and 100 random images from each of 34 adult monkeys (30 female and 4 male). The randomly selected images included faces that were partially obscured, rotated or poorly lit. The images were collected from videos taken on multiple days over a one year period. The well-aligned images were used to train the different face recognition systems, both well-aligned and random images were used to test the different algorithms.

It is possible to train a face recogniser using the output of the face detection system but this would have a low success rate and be very susceptible to changes in lighting and pose. A number of methods for processing the faces to overcome this including EigenFaces (based on principal component analysis) and FisherFaces (based on linear discriminant analysis)[5]. A method based on local binary patterns[6] has been shown to be relatively robust to changes in light intensity, which is important for this study as the enclosures are lit by a mixture of natural and artificial light. There are a number of different local binary patterns and I have used the most basic. This involves comparing the intensity of each pixel with its 8 surrounding pixels (shown in inset in Figure 2) and assigning a 1 if the central pixel is lighter than the surrounding pixel and 0 if it is darker. An 8-bit binary code is then assigned to the central pixel (this relates to a number between 0 and 255). Some patterns are more informative for face recognition than others. A uniform local binary pattern is one where there are only a couple of changes from 0 to 1 (or vice versa) in the pattern and usually denotes a feature such as an edge or corner[6]. Throughout this project I have only considered uniform patterns. The conversion of the face to local binary patterns is one of a number of processing stages required before the faces can be classified (stages illustrated in Figure 2).

After initial face detection, each face is resized to 100 x 100 pixels and converted to grayscale (Figure 2A). The alignment of the face is checked as shown in Figure 1B. A mask is applied to focus on the central portion of the face (Figure 2C) and the image is converted to local binary patterns (Figure 2D and inset). The face is then divided up into 25 equally sized blocks (Figure 2E). For each block a histogram is formed of the uniform local binary patterns (Figure 2F) and then the histograms from each block are concatentaed to form one long histogram (Figure 2G). These histograms form the basis for the classification stage.

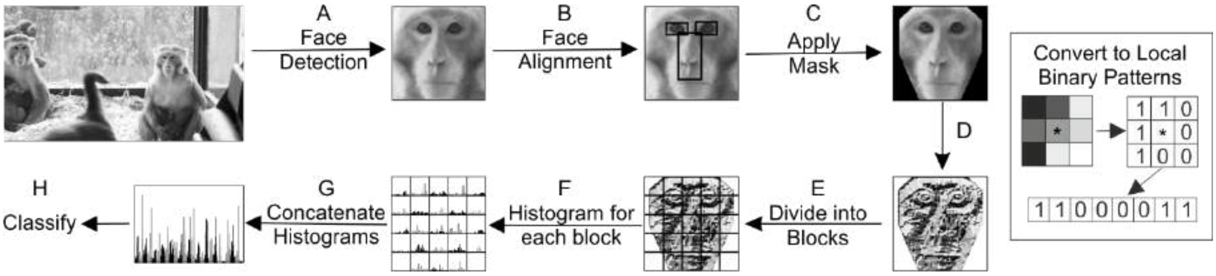


Figure 2. Diagram of face recognition stages. Inset, calculating the local binary pattern for a single pixel (*).

I tested the success rate of three different maching learning algorithms –nearest neighbour (NN)[6], local discriminant analysis (LDA)[7] and support vector machines (SVM; using a one *vs* one approach for mulitple classes)[8]. For the LDA classifier the dimensionality of the local binary pattern histogram was first reduced using principal component analysis[7]. Each algorithm was implemented using the standard Matlab functions. For the testing each classifier was trained on a subset of the monkeys and well-aligned images. M individuals were randomly selected from the total of 34 monkeys (M ranged from 2 to 32) and N well-aligned face images randomly selected for each monkey (N ranged from 2 to 40). The classifier was then trained using the local binary pattern histograms as illustrated in Figure 2H. The success rate was calculated by using the trained classifier to predict the identity of a further 10 well-aligned images per individual (for a total of 10*M images per parameter combination). This was then repeated 10 times for each parameter combination and the average success rate taken. A subset of these results is shown in Figure 3A. For very small numbers of monkeys (e.g. pairs) the success rate of the three classifiers was comparable (Figure 3A, black bars; success rate shown for 20 images per individual). As the number of monkeys increased the success rate of all three algorithms decreased but the LDA algorithm consistently produced the highest success rate under the different conditions (Figure 3A, white bars; 32 monkeys with 20 images per individual used for training).

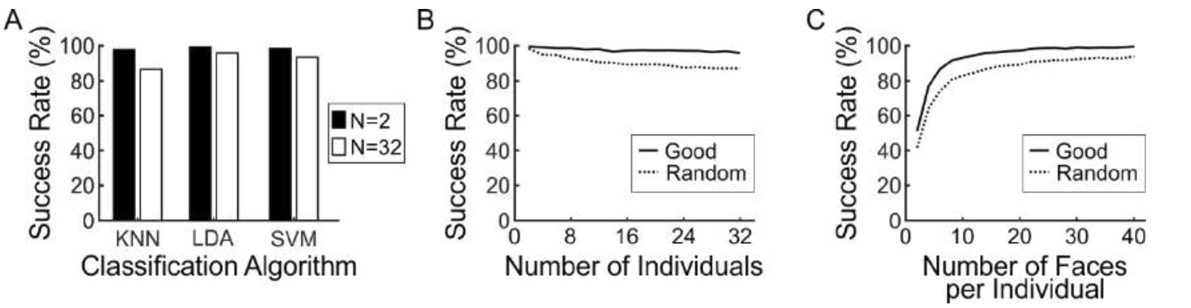


Figure 3. Success of face recognition. A, effect of classification algorithm on success rate. B, effect of increasing number of individual monkeys on success rate. C, effect of increasing number of sample faces per monkey in the training set on success rate.

As the LDA approach gave the highest success rates I then used that approach to assess the effects of varying the number of individuals and the number of images per individual on the successful classification of the well-aligned (“Good” in Figure 3B-C) and random (“Random” in Figure 3B-C) faces. For the well-aligned images the success rate remained >95% as the number of monkeys increased from 2 to 32 (Figure 3B, solid line; 20 images per individual used for training). For the randomly selected images, which included the images that were partially obscured, rotated or poorly lit, the success rate dropped to 85% for 32 monkeys (Figure 3B, dotted line). The number of images per individual used for training had a large effect on the success rates of both well-aligned images and random images (Figure 3C; 16 individuals used for training). Low numbers of images (less than 8 images per individual) were particularly poor. As the numbers of images per individual increased the success rate improved more slowly and between 20 and 40 images there was only a small improvement in success rate. In a further test I looked at different conditions that might affect the success rate of the face recognition – high contrast images due to sunlight (“high contrast”), poorly aligned images due to rotation of head (“rotation”), partially obscured images (“obscured”) and images with facial expressions (“expression”). Of these four categories “rotation” had the most detrimental effect, reducing the success rate to 65% for 32 monkeys. “High contrast” also had a detrimental effect on success rate (73% for 32 monkeys). “Expression” and “obscured” images had similar success rates to the randomly selected images above (89% and 81% respectively).

Finally I tested a possible application of face recognition to animal social analysis. The face recognition algorithms were run on three days of video of one group of 10 female macaques. A pair of females were considered associated if there were at least three detections of each female within a 1 minute window. I then compiled the association matrix using the free software Socprog [10]. The strongest associations between the 10 females are shown in Figure 4, with the three pairs of full sisters in the group indicated by circles. To validate this application the videos were also scored using traditional coding methods (scan sampling at 1 minute intervals) and a second association matrix produced based on the manually scored data. The association matrix produced by the automated analysis was significantly correlated with the association matrix from manual analysis (Mantel Test; matrix correlation of 0.723; P<0.001; analysed using Socprog).

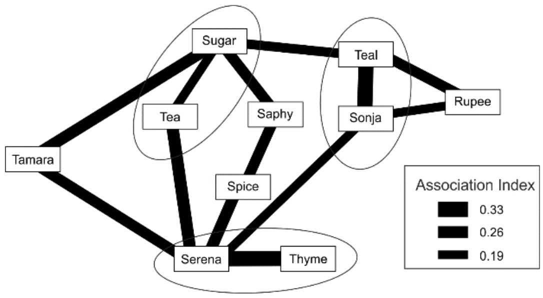


Figure 4. Face recognition and social analysis. Group social structure based on automated analysis between 10 adult female macaques. Line thickness indicates strength of association and pairs of full sisters are identified by circles.

Conclusions

I have shown, for the first time, that it is possible to use automated face recognition to identify individual rhesus macaques with a success rate of over 95% for well-aligned images. Including images which may be rotated, obscured or poorly lit reduced the success rate to 85%. In future either these type of images should be automatically excluded or the face recogniser made more robust. A similar study in chimpanzees encountered the same problems[9]. Not surprisingly the face recogniser performed best when trained on small groups of monkeys and high numbers of images per individual. However between 10 and 20 images per monkey produced reasonable success rates for up to 32 monkeys and I have found whilst collecting the videos for this project that 1-2 days recording is sufficient to provide 20 images per individual for all monkeys within a group. I have demonstrated a potential application of the face recognition to social network analysis based on proximity between animals. In the future it may be possible to combine this face recognition approach with automated behaviour monitoring.

References

1. Kühl, H.S. and T. Burghardt, *Animal biometrics: quantifying and detecting phenotypic appearance.* Trends in Ecology & Evolution. **28**(7): p. 432-441.
2. Parr, L.A., et al., *Recognizing Facial Cues: Individual Discrimination by Chimpanzees (Pan troglodytes) and Rhesus Monkeys (Macaca mulatta).* Journal of comparative psychology (Washington, D.C. : 1983), 2000. **114**(1): p. 47-60.
3. Loos, A. and A. Ernst, *An automated chimpanzee identification system using face detection and recognition.* EURASIP Journal on Image and Video Processing, 2013. **2013**(1): p. 1-17.
4. Viola, P. and M. Jones. *Rapid object detection using a boosted cascade of simple features*. in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. 2001.

5. Belhumeur, P.N., J.P. Hespanha, and D. Kriegman, *Eigenfaces vs. Fisherfaces: recognition using class specific linear projection.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. **19**(7): p. 711-720.
6. Ahonen, T., A. Hadid, and M. Pietikainen, *Face description with local binary patterns: application to face recognition.* IEEE Trans Pattern Anal Mach Intell, 2006. **28**(12): p. 2037-41.
7. Fratric, I. and S. Ribaric, *Local Binary LDA for Face Recognition*, in *Biometrics and ID Management: COST 2101 European Workshop, BioID 2011, Brandenburg (Havel), Germany, March 8-10, 2011. Proceedings*, C. Vielhauer, et al., Editors. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 144-155.
8. Lian, H.-C. and B.-L. Lu, *Multi-view Gender Classification Using Local Binary Patterns and Support Vector Machines*, in *Advances in Neural Networks - ISNN 2006: Third International Symposium on Neural Networks, Chengdu, China, May 28 - June 1, 2006, Proceedings, Part II*, J. Wang, et al., Editors. 2006, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 202-209.
9. Loos, A. *Identification of great apes using gabor features and locality preserving projections*. in *1st ACM international workshop on Multimedia analysis for ecological data*. 2012. ACM: Nara, Japan.
10. Whitehead, H. 2009. SOCPROG programs: analyzing animal social structures. Behavioural Ecology and Sociobiology. 63: 765-778.

Neurophysiological study of attentional and working memory processes in rats: focus on late P300 event-related potentials in the active oddball paradigm

L. Raeymaekers, A. Ahnaou, R. Biermans, J.A. Kemp and W.H. Drinkenburg

Dept. of Neuroscience, Janssen Research & Development, a Division of Janssen Pharmaceutica, Beerse, Belgium. lraeymae@its.jnj.com

Abstract

Cognitive event related potentials (ERPs) and synchronicity of evoked oscillations both provide powerful read-outs for studying the synaptic functions of the brain that are underlying information processing in real time. The P300 component of cortical ERPs, indexing attention and working memory, is altered in neurological and psychiatric diseases. Structures of the limbic system are strongly involved in the generation of P300, and pharmacological manipulation of acetylcholine and related receptors may affect cognitive performance and P300 response [1,2]. The present study investigated: 1) whether the P300 response, associated with adequate functioning of underlying neuronal circuits and evoked oscillations synchrony, could be generated in rats performing in an auditory discrimination task, and 2) to what extent modulation of cholinergic neurotransmission, through administration of scopolamine and donepezil, contributes to changes in P300 components associated with cognitive performance.

All experimental protocols were carried out in strict accordance with guidelines of the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC), and of the European Communities Council Directive of 24th November 1986 (86/609/EEC) and were approved by the ethical committee of Janssen Research and Development. Behavioral training and electroencephalographic (EEG) recording sessions took place in a standard operant chamber (MED Associates, Inc.). Task-relevant auditory evoked potentials (AEP) components were recorded in multiple cortical areas of male Long Evans rats performing an auditory discrimination task, comparable to the "odd-ball" paradigm used in human experiments. Rats were trained to discriminate two auditory stimuli, a frequent non-target tone and a rare target tone related to food reward during baseline conditions and following modulation of the cholinergic tone with scopolamine (0.64 mg/kg, sc) and donepezil (1 mg/kg, sc)[3].

Rats that consistently discriminate target from non-target tones showed P300 peaks with early and late components, whereas this potential was not clear for animals that did not discriminate relevant stimuli. Donepezil enhanced P300 amplitude and evoked theta/gamma oscillations to target tones and attenuated scopolamine-induced deficits in premature behavioral responses and AEPs/synchronicity of evoked oscillations.

This study provides important evidence for changes of P300-like potentials in rats, who are engaged in behavioural task with a demand on both attention and working memory. Furthermore, findings highlight the relevance of the cholinergic system in stimulus discrimination processing. The rodent AEP P300 paradigm shows added value as a translational tool to assess targets and/or pharmacological agents that contribute to the facilitation and suppression of synaptic responses underlying early and late P300 AEP components.

References

1. Laursen B, Mørk A, Kristiansen U, Bastlund JF. Hippocampal P3-like auditory event-related potentials are disrupted in a rat model of cholinergic degeneration in Alzheimer's disease: reversal by donepezil treatment. *J Alzheimers Dis.* 2014; **42**:1179-89.
2. Chang YS, Chen HL, Hsu CY, Tang SH, Liu CK. Parallel improvement of cognitive functions and P300 latency following donepezil treatment in patients with Alzheimer's disease: a case-control study. *J Clin Neurophysiol.* 2014; **31**:81-5.

3. Ahnaou A, Huysmans H, Jacobs T, Drinkenburg WH. Cortical EEG oscillations and network connectivity as efficacy indices for assessing drugs with cognition enhancing potential. *Neuropharmacology*. 2014; 86:362-77.

Quantitative Evaluation of Oromotor Function: A Simple and Stress-free Test for Motor Behavior

R. Rablm, A. Horvath, and B. Hutter-Paier

QPS Austria GmbH, Neuropharmacology, Grambach, Austria,
roland.rabl@qps.com

Measuring motor behavior in rodent disease models is often performed in long lasting tests with time consuming evaluations (paper pencil) under high stress conditions (e.g. Rota Rod, Challenging Beam Walk). Especially testing of rodent models with a severe phenotype or increased probability of epileptic seizures sometimes leads to insufficient results due to an incapability of the animals to perform the task or even death incidences after seizures. For this reason, we wanted to develop a motor test for rodents with a minimum of experimenter`s or equipment interference and thus a minimum of stress for the animals.

In 2011, Kane et al. noticed that when performing a Pasta handling test (Vermicelli and Capellini handling test) for evaluating the lesion rate of unilaterally 6-OHDA lesioned rats, not only the number of adjustments with each paw was altered, but also the gnawing noise was different. Since eating displays a basal natural behavior and gnawing noise can also be easily recorded from a distance, we decided to establish quantitative analysis of this behavior.

Various neurodegenerative diseases like Parkinson`s disease, Amyotrophic lateral sclerosis and Niemann-Pick disease are accompanied by speech and swallowing disturbances or mask-like face expressions in humans, so we checked whether this finding could also be observed in motor impaired transgenic mice. Therefore, we evaluated the gnawing noise of three mouse models of three different indications known to present motor deficits.

Mouse models of Parkinsons disease (TNWT#61 – Line61), Amyotrophic lateral sclerosis (Tar6/6 – TDP43x TDP43), Niemann-Pick disease (NPC1 ko/ko) and corresponding wild type (WT) mice at various ages have been tested. Additionally, animals underwent two classical motor tests (Rota Rod and Wire Hanging Test). Animal experiments were approved by the Styrian government (Austria) and performed under the Austrian guidelines for the care and use of laboratory animals. Housing of animals and behavioral experiments were done in a fully AAALAC accredited facility.

Pasta gnawing protocol:

Two hours prior to testing the food pellets of all animals were removed. One little piece of dry spaghetti was given to each animal to become familiar with the novel food.

To measure gnawing noise, we placed a microphone above the home cage and put dry spaghetti pieces (approx. 1 cm long) into the home cage. Afterwards, we recorded the gnawing noise. Acquisition was performed by using. Behringer ECM 8000 microphone connected to a Steinberg Cl1 audio interface. Steinberg Wave Lab LE 7 was used as recording software. The acquired gnawing pattern was analyzed using sound analyzing software (Avisoft SASLab Pro 5.1).

Two parameters were evaluated:

- Biting speed (gnawing interval length)
- Number of bites during a chewing period (gnawing peaks per episode).

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

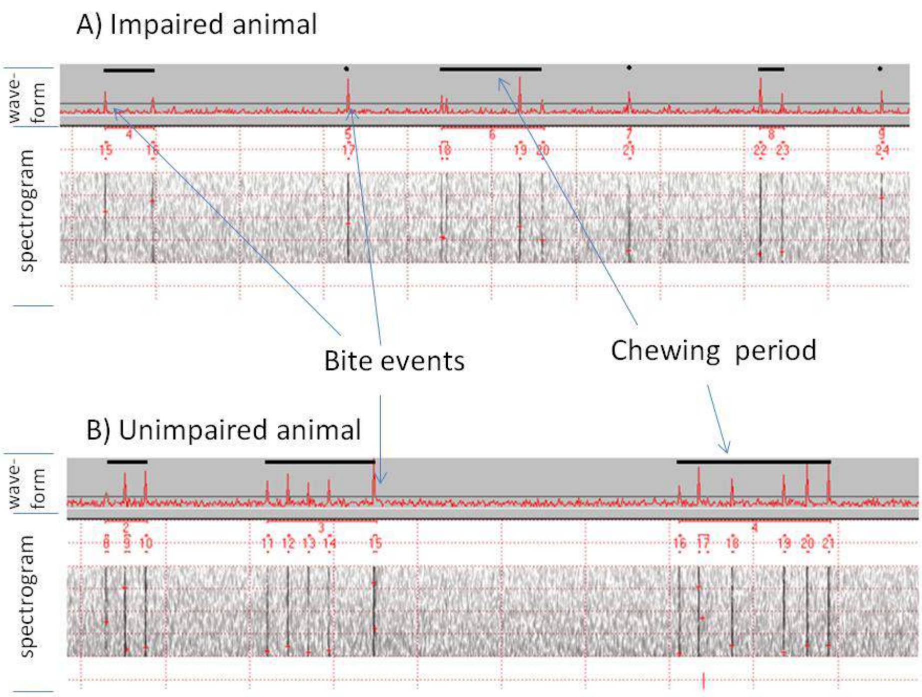


Figure 1: Exemplary gnawing pattern of an impaired (A) and unimpaired (B) animal.

Results:

TNWT#61:

TNWT#61 mice showed progressive motor impairments starting at 3 months of age compared to wt littermates in terms of gnawing peaks per episode (B) in the Pasta Gnawing Test. Highly significant differences were also observed at all ages in the Rota Rod (C) and wire hanging performance (D). Evaluation of the biting speed (A) did not reach significance, but showed a trend of decreasing gnawing speed in TNWT#61 mice.

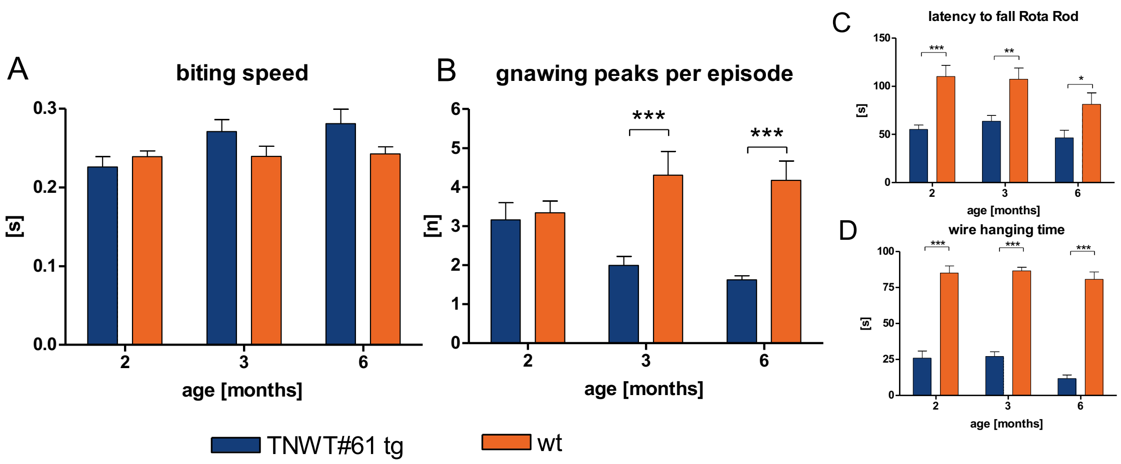


Figure 2. Motor evaluation of TNWT#61 animals. Biting speed (A) and gnawing peaks per episode (B) of TNWT#61 male animals compared to wildtype (wt) littermates as well as latency to fall off the Rota Rod (C) and wire hanging time (D) of the

same animals. n=10-15 per group and age; Statistical analysis: Two-way-ANOVA followed by Bonferroni post-test, * p<0.05; **p<0.01; ***p<0.001.

TDP43xTDP43:

TDP43xTDP43 mice showed progressive motor impairments compared to wt littermates in terms of gnawing peaks per episode (B) in the Pasta Gnawing Test similar to the Rota Rod (C). Evaluation of the Wire hanging performance (D) resulted in highly significant differences at both ages. Furthermore, a significant decrease of biting speed (A) could be detected at 20 weeks of age.

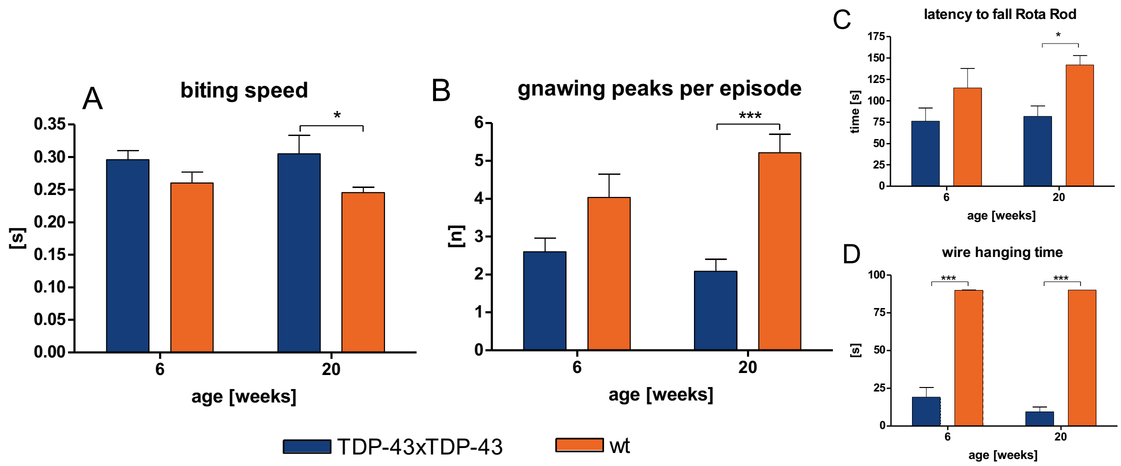


Figure 3. Motor evaluation of TDP43xTDP43 animals. Biting speed (A) and gnawing peaks per episode (B) of TDP43xTDP43 animals compared to wt littermates as well as latency to fall off the Rota Rod (C) and wire hanging time (D) of the same animals. n=5-16 per group and age; Statistical analysis: Two-way-ANOVA followed by Bonferroni post-test, * p<0.05; ***p<0.001.

NPC1 ko/ko:

NPC1 ko mice showed motor impairments compared to wt littermates in terms of gnawing peaks per episode at an age of 8 weeks (B) in the Pasta Gnawing Test similar to the Rota Rod (C). Gnawing speed (A) and wire hanging duration (D) were not affected by the genotype.

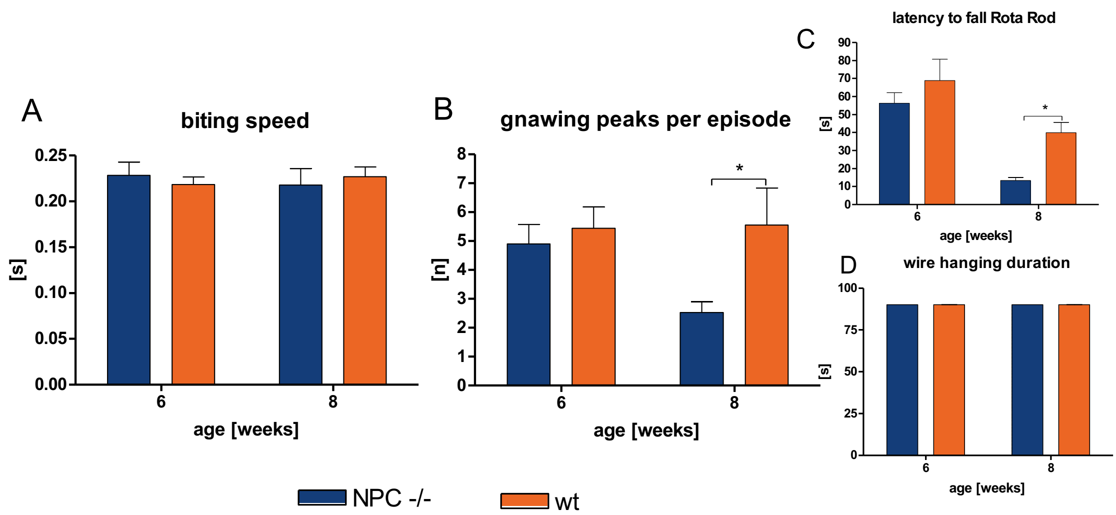


Figure 4. Motor evaluation of NPC1 ko animals. Biting spped (A) and gnawing peaks per episode (B) of NPC1 ko animals compared to wt littermates as well as latency to fall off the Rota Rod (C) and wire hanging time (D) of the same animals. n=9-12 per group and age; Statistical analysis: Two-way-ANOVA followed by Bonferroni post-test, * $p<0.05$;

Conclusion

Our results show, that the Pasta gnawing test is capable to quantitatively detect motor disturbances in mouse models of various neurodegenerative diseases similar to conventional motor tests. We furthermore were able to detect a more robust sensitivity towards disease progression compared to Rota Rod and Wire Hanging Test.

As a conclusion, the Pasta Gnawing Test displays a novel, powerful tool to characterize mouse models of neurodegenerative diseases, as well as for efficacy tests investigating new compounds. Simultaneously, the test is especially sensitive resulting in more robust results and also more importantly, reducing stress caused by animal handling compared to use of other motor tests.

References

1. Kane, J.R., Ciucci, M.R., Jacobs, A.N., Tews, N., Russel, J.A., Ahrens A.M, Ma, S.T, Britt, J.M., Cormack, L.K., Schallert, T. Assessing the role of dopamine in limb and cranial-oromotor control in a rat model of Parkinson`s disease. *J CommunDisord.* 2011 Sep –oct;44(5):529-37. Doi:10.1016/j.jcomdis.2011.04.0050 Epub 2011 Apr 30.

Prefrontal Circuits in Fear Discrimination Learning and Psychiatric Vulnerability

J.W. Lovelace, P.A. Vieira, E. and E. Korzus

Department of Psychology, University of California, Riverside, CA, USA.
edkorzus@ucr.edu

Abstract

We have recently investigated prefrontal mechanisms underlying fear discrimination learning. The fear discrimination task examines the ability of mice to distinguish aversive and harmless stimuli (1, 2). After being handled, individual mice were exposed to context A one day before training. The protocol included 14 days of training, which was divided into three phases: initial training phase, generalization test and discrimination phase. During the initial training phase (day 1), mice were placed in the context A for 180 s followed by a single foot shock (arrow) and left for another 60 s inside the chamber. Context A (CS+) was the unmodified fear conditioning box (Coulburn Instruments Inc.), which was placed inside of a sound attenuated chamber with the house light and house fan on. For generalization test and during discrimination phase, the individual mice were exposed to Context A for 180 s and received a 0.75 mA, 2 s foot shock, and left for another 60 s inside the chamber. Four hours later, the mice were exposed to the similar Context B (CS-) for 242 s and received no footshock. Context A and B were similar but not the same. Context B was the modified fear conditioning chamber, with angular wall inserts, house fan off, and scented with Simple Green. Thus animals were exposed to CS^{+} 13 times before the final test. The order of exposure to different contexts was counter balanced. Additionally, the context cues themselves were counter balanced within each group in order to isolate the effect of the CS+.

To investigate the role of epigenetic regulation in fear discrimination learning, we generated a CREB Binding Protein (CBP) mutant lacking HAT activity and introduced the mutated gene into transgenic mice to specifically block the HAT activity of CBP in a subset of neurons in the living adult brain. Our current research investigates how the mPFC-amygdalar-hippocampal circuit acts via multiple pathways to exert both excitatory and inhibitory influences on fear responses under the central hypothesis that the accuracy of fear memory is attained via the mPFC-dependent decline of fear responses to harmless stimuli. This hypothesis is partially based on our strong data demonstrating that fear discrimination learning is disrupted by the acute hypofunction of mediators of cellular and psychological memory consolidation [such as CREB or CBP's HAT] in the mPFC (2). These studies suggest that the mPFC is required for a reduction of generalized fear to harmless stimuli that is essential for improvement of fear memory accuracy. In addition, these data indicate that certain types of prefrontal dysfunction likely contribute to overgeneralized fear, a clinical condition present in anxiety-related disorders such as PTSD.

Our research indicates that abnormal interactions of neurodevelopment with the environment triggered by drugs of abuse during neonatal (1) or adolescent (3) periods may permanently impair brain function, such as endocannabinoid system (eCB)-dependent inherent neuroprotection of circuit integrity and neuroplasticity. Deficiency in eCB signaling was associated with abnormalities in fear discrimination learning and fear extinction. We show that mouse model of adolescent cannabis abuse shows deficits in an endocannabinoid-mediated signaling and neuroplasticity in adult prefrontal cortex, a brain region encompassing neural circuit for decision-making. The eCB system represents a major activity-dependent regulatory system in the central nervous system and has been implicated in multiple brain functions, including synaptic plasticity and the homeostatic regulation of network activity patterns. Whereas current data link cannabis abuse in adolescence to increased risk for dependence on other drugs, depression, anxiety disorders (~18% of Americans) and psychosis (~1% of Americans), the mechanism(s) underlying these adverse effects remains controversial. Despite the importance of

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

this knowledge, the adolescent ontogeny of the eCB system along with its effects on network homeostasis remains elusive.

Subjects: The UC Riverside Institutional Animal Care and Use Committee approved all procedures in accordance with the NIH guidelines for the care and use of laboratory animals. We used C57BL/6J mice for all experiments. Mice were weaned at postnatal day 21, housed 4 animals to a cage with same sex littermates with ad libitum access to food and water and maintained on a 12 h light/dark cycle. Old bedding was exchanged for fresh autoclaved bedding every week.

References:

1. Lovelace JW, Vieira PA, Corches A, Mackie K, & Korzus E (2014) Impaired fear memory specificity associated with deficient endocannabinoid-dependent long-term plasticity. *Neuropsychopharmacology* 39(7):1685-1693.
2. Vieira PA*, et al.* (2014) Prefrontal consolidation supports the attainment of fear memory accuracy. *Learn Mem* 21(8):394-405.
3. Lovelace JW*, et al.* (2015) An animal model of female adolescent cannabinoid exposure elicits a long-lasting deficit in presynaptic long-term plasticity. *Neuropharmacology* 99:242-255.

Influence of Density and Environmental Constraints on the Shoaling Behavior of Two Species of Fish

E. Gimeno, V. Quera, F. S. Beltran, and R. Dolado

Department of Behavioral Science Methods, University of Barcelona, Barcelona, Spain.
egimeno@ub.edu

Introduction

Many fish species live in social groups, most likely because they can achieve significant anti-predatory and foraging advantages. Social groups of fish display several types of collective motion, whose features can change in response to external stimuli such as predators or food. The most common types of collective motion of fish are shoals, a group of fish that has a significant degree of cohesion, and schools, a group of fish that has a significant degree of both velocity and alignment synchronization. Other kinds of collective behavior have been defined [1]. There are not many comparative studies looking for differences in the collective motion of different species under laboratory conditions [2-3].

The collective motion of fish is usually studied by multitracking groups of several individuals that are video recorded, but the experimental procedures and variables used can vary greatly from one researcher to another. Some authors use a shallow water column, about 5 cm [2] to prevent fish to superimpose on other individuals as much as possible by reducing their vertical mobility, whereas others use a water depth up to 30 cm [4]. A second difference is the geometry of the tank used, which can be circular [4], rectangular [5] or square-shaped [6]. Moreover, researchers use many different group sizes, ranging from 4 [5] to 50 fish [7].

The aim of this study was to compare the collective motion of two fish species, zebrafish (*Danio rerio*) and black neon tetras (*Hyphessobrycon herbertaxelrodi*). The zebrafish is a small fish of the Cyprinidae family living in slow-flowing waters like pools, rice paddies and stream margins in Pakistan, India and Bangladesh [8]. These habitats have abundant vegetation, numerous places for fish to hide, and are probably bereft of open areas, which make fish groups little conspicuous to predators. Moreover, zebrafish tends to be vertically distributed throughout the whole of the water column [8]. Hyphessobrycon is one of the largest genera in the family Characidae and comprises about 130 species distributed from Southern Mexico to the Río de la Plata in Argentina [9]. The black neon tetra (*H. herbertaxelrodi*) is a well-known species in the aquarist business, but as far as we know its behavior in the wild has been received minimal study, so little is known of its ecology, group size, vertical distribution when swimming, and aquatic and avian predators.

We compared the two species because they are supposed to show different types of collective motion according to their different tendency to superimpose and occlude while they swim in groups of conspecifics [10]. On the other hand, we systematically explored the effect of group size (n = 10 and n = 20) and environmental constraints, such as water column height (15 and 25 cm) and tank geometry (rectangular and rounded corners), to test whether the use of different experimental procedures leaded to significant changes in the collective motion of the fish. Furthermore, we characterized the collective motion of the black neon tetra, whose collective motion has rarely been studied.

Material and methods

We recorded the movement of groups of each fish species (n = 10 and 20) during 30 min, using two tank geometries: a rectangular shape and a rectangular shape with rounded corners (see Figure 1); and we also manipulated the water height (15 and 25 cm). We randomly selected six video fragments of 25 s per experimental condition, and applied a method to resolve occlusions described elsewhere [10]. Those occlusions that were not automatically resolved by the method were manually split by a trained observer using ImageJ (NIH open-source software) to obtain a stack of occlusion-free frames. Finally, we acquired the individual fish

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

trajectories for each fragment using Image-Pro Premier® 9.1 in the occlusion-free stacks. We employed the individual trajectories to calculate indexes of cohesion, coordination, and shoal shape, such as mean interindividual distance (IID), coefficient of variation of the mean interindividual distances (CV_{IID}), shoal density, global polarization (i.e., the tendency of individuals to adopt the same orientation as its group mates [1]), and circularity. To calculate the global polarization we used a similar approach as the Cohen's kappa coefficient [11]: we compared the observed polarization, $\rho(t)$, with the expected one under the hypothesis of random distribution of individual movement vectors, E[ρ], by calculating the ratio of the difference between the observed and expected polarization to the maximum possible difference. We named this measure kappa-polarization (κ_ρ).

$$\kappa_\rho(t) = \frac{\rho(t) - E[\rho]}{1 - E[\rho]}$$

The mathematical expectancy of polarization, E[ρ], was estimated by creating n (10 or 20) random movement vectors repeatedly and independently 20,000 times, calculating ρ for each one and averaging the ρs.

Results and discussion

We found that the two species showed different patterns while swimming in a group. Figure 2 summarizes the results of the collective motion of the fish in all the experimental conditions.

An ANOVA showed a significant quadruple interaction between water column height × group size × geometry × species on the polarization of the fish, as measured by κ_ρ ($F(1,80) = 5.2$, $p < .001$). Post-hoc pairwise comparisons revealed that zebrafish had a lower polarization than black neon tetras in all the experimental conditions. Black neon tetras were highly coordinated regardless of water column height, tank geometry or group size, whereas polarization in zebrafish was sensitive to tank geometry and water column height when group size was small.

Two significant interactions were found between water column height × group size × species ($F(1,80) = 11.55$, $p = .001$) and water column height × geometry × species ($F(1,80) = 34.20$, $p < .001$). Zebrafish tended to swim significantly farther away from each other more frequently than black neon tetras in almost all cases. Both species increased their IID when group size was larger, except for zebrafish in a deep water column. We also found a significant quadruple interaction between water column height × group size × geometry × species on the CV_{IID} ($F(1,80) = 4.41$, $p = .039$). Zebrafish broadly tended to scatter more than black neon tetras, which suggests that some individuals were farther away from the rest of the group or even that the group split into several subgroups.

We found a quadruple interaction between water column height × group size × geometry × species on the density of the fish ($F(1,80) = 19.11$, $p < .001$). The density of black neon tetras was higher than that of zebrafish, which

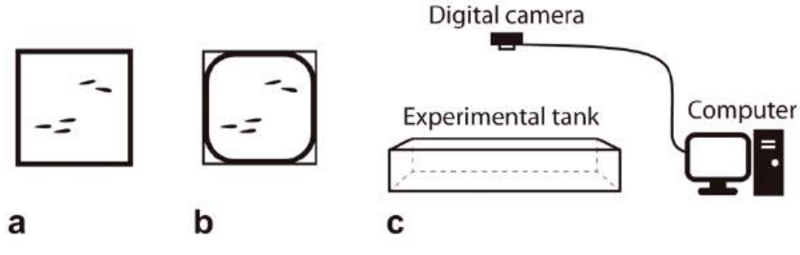


Figure 1. Schematic representation of the experimental setup. a) Top view of the tank with a rectangular geometry. b) Top view of the tank with rounded-corner geometry. c) A digital camera above the experimental tank sends a digital signal to a computer, which records the videos.

is consistent with the IID results. If fish tend to be distributed vertically, we expect to find differences in their cohesion and density when water column height increases because, when seen from above, individuals arranged upward and downward appear closer in the horizontal plane. While the results for zebrafish shoals matches this hypothesis, the IID for black neon tetra schools did not show any difference based on water column height and the results on density did not fit the expected pattern. This suggests that the two species do not make the same use of vertical space.

Finally, an interaction was found between water column height × geometry × species on the circularity of the shoal shape ($F(1,80) = 5.62$, $p = .020$). Zebrafish shoals tended to be less circular than black neon tetra shoals. When the water column was deep, the zebrafish shoals had a more circular shape in the rounded-corner tank than in the rectangular tank, whereas when they swam in a rectangular tank, shoals were more circular when the water column was shallow.

To conclude, this study shows that black neon tetras are schooling species that display very consistent polarization even with increased group size; their groups remain compact and their tendency to split is much lower than in shoaling species such as the zebrafish. Therefore, black neon tetra may be a suitable model to study the collective motion of schooling species, since it is a small organism that is easy and cheap to keep in large groups. Black neon tetras did not seem to distribute vertically in the water column as much as zebrafish did, at least when swimming in a group. We also found that group size had an effect on cohesion and density in the two species, whereas environmental constraints such as water column height and tank geometry did not have a robust effect on group polarization, cohesion, density or shape. However, shoaling species appear to be more sensitive to such constraints than schooling species, which suggests that the collective motion of schools is more robust than the motion of shoals.

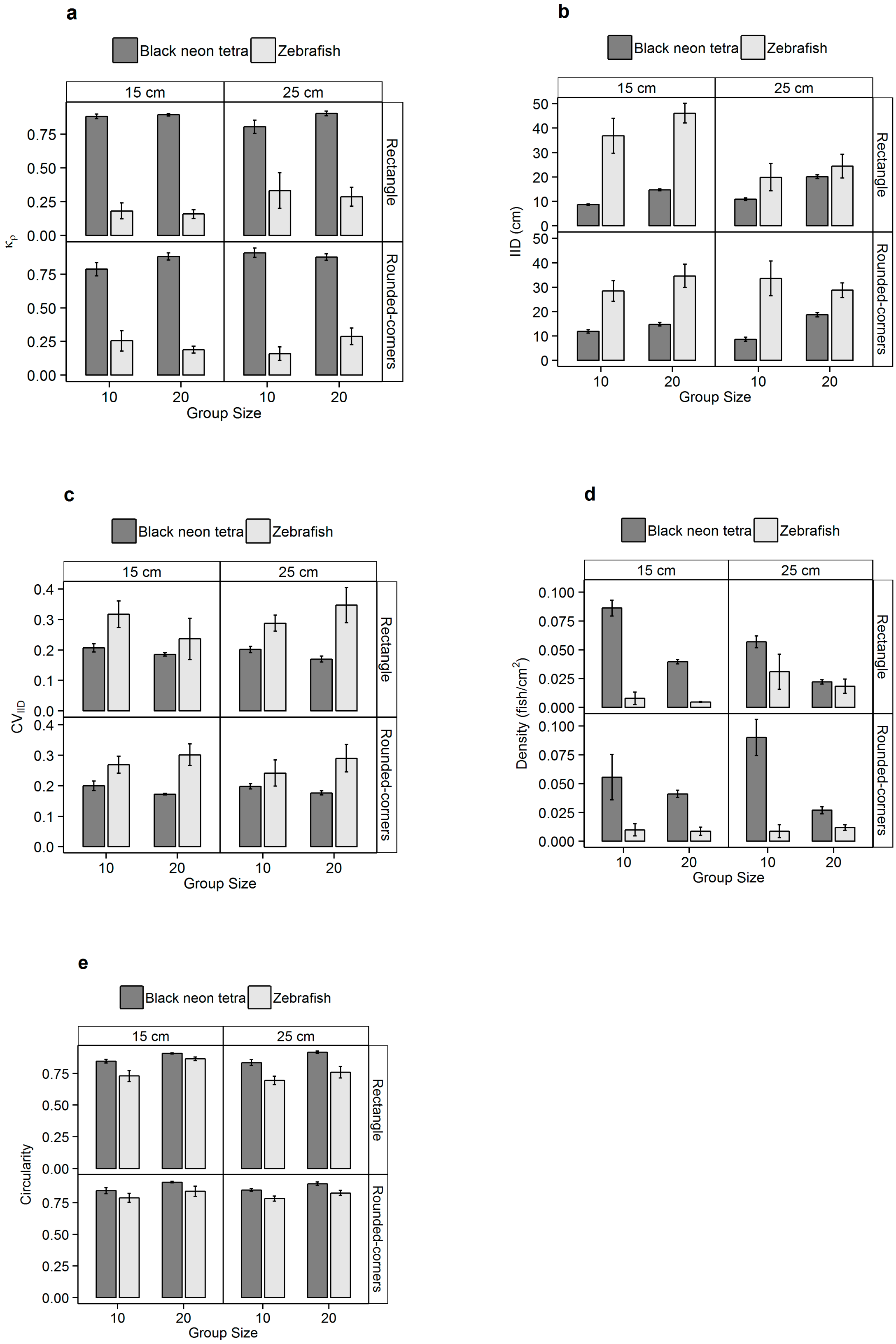


Figure 2. Mean values, averaged for each experimental condition, of a) *κρ*, b) IID, c) CV_{IID}, d) Density, and e) Circularity. Error bars denote 95% CI.

References

1. Delcourt, J., Poncin, P. (2013). Shoals and schools: Back to the heuristic definitions and quantitative references. *Reviews in Fish Biology and Fisheries*, **22**, 595–619.
2. Leem, J. B., Jeon, W., Yun, C. Y., Lee, S. H. (2012). Quantitative analysis of fish schooling behavior with different numbers of medaka (*Oryzias latipes*) and goldfish (*Carassius auratus*). *Ocean Science Journal*, **47**, 445–451.
3. Soria, M., Fréon, P., Chabanet, P. (2007). Schooling properties of an obligate and a facultative fish species. *Journal of Fish Biology*, **71**, 1257–1269.
4. Miller, N., Gerlai, R. (2007). Quantification of shoaling behaviour in zebrafish (*Danio rerio*). *Behavioural Brain Research*, **184**, 157–166.
5. Shelton, D.S., Price, B.C., Ocasio, K.M., Martins, E.P. (2015). Density and group size influence shoal cohesion, but not coordination in zebrafish (*Danio rerio*). *Journal of Comparative Psychology*, **129**, 72–77.
6. Salierno, J.D., Gipson, G.T., Kane, A.S. (2008). Quantitative movement analysis of social behavior in mummichog, *Fundulus heteroclitus*. *Journal of Ethology*, **26**, 35–42.
7. Miller, N., Gerlai, R. (2012). From schooling to shoaling: Patterns of collective motion in zebrafish (*Danio rerio*). *PLOS ONE*, **7**, e48865.
8. Spence, R., Fatema, M.K., Reichard, M., Huq, K.A., Wahab, M.A., Ahmed, Z.F., Smith, C. (2006). The distribution and habitat preferences of the zebrafish in Bangladesh. *Journal of Fish Biology*, **69**, 1435–1448.
9. Carvalho, F.R., Langeani, F. (2013). *Hyphessobrycon uaiso*: New characid fish from the rio Grande, upper rio Paraná basin, Minas Gerais State (Ostariophysi: Characidae), with a brief comment about some types of Hyphessobrycon. *Neotropical Ichthyology*, **11**, 525–536.
10. Dolado, R., Gimeno, E., Beltran, F.S., Quera, V., Pertusa, J.F. (2015). A method for resolving occlusions when multitracking individuals in a shoal. *Behavior Research Methods*, **47**, 1032–1043.
11. Cohen, J.A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

Social Representations in the Mouse Prefrontal Cortex

Levy, D.R. [1], Weissbrod, A., [1] Yizhar, O [1].

[1] Dept. of Neurobiology, Weizmann Institute of Science, Rehovot, Israel. Dana-rubi.Levy@weizmann.ac.il

The medial prefrontal cortex (mPFC) has been widely implicated in orchestrating a variety of social functions in the mammalian brain, including active modulation of complex social behaviors [1-4]. Abnormalities in the activity and morphology of this region are also often associated with psychiatric disorders that involve social deficits such as schizophrenia and autism [5-7]. Despite the accumulating evidence for the role of the mPFC in modulation of sociability, it is still unclear how, or even whether, the mPFC encodes specific social information. Furthermore, the extent to which such social representations might be involved in regulating social behavioral responses, in health and disease, remains unknown. To address these questions, we recorded stimulus-evoked neuronal activity in the mPFC of freely behaving mice presented with a repertoire of highly-controlled social and non-social sensory cues. We then systematically characterize the dynamics of mPFC unit responses to the various information carried by these stimuli, as well as to the behavioral responses which they evoke.

Recording the neural representations of social information necessitates an apparatus that allows the precisely-timed delivery of various sensory stimuli. We therefore designed a novel apparatus which enables in vivo recording in behaving mice coupled with rapid and tightly-regulated presentation of both olfactory and visual cues (Fig. 1a). The setup is comprised of two transparent cages holding the experiment and stimulus mice. The cages are separated by an LCD shutter designed to shift between transparent and opaque states in response to change in voltage, thus exposing the visual stimulus to the experiment mouse (Fig. 1b). Measurements of light transmission through the shutter confirm that it alternates between states within ~4 ms (Fig. 1c). Olfactory stimuli are carried to the experimental cage by a custom made olfactometer, which provides a constant flow of air delivering the selected odors (Fig. 1a). A vacuum pump continuously clears air from the cage in order to maintain air flow and remove odor residue throughout the experiment. Calibration of odor concentration in the apparatus was performed using a a photoionization detector (PID) and pressure sensors. We found that odors reach the experimental cage within ~20 ms of cue delivery (Fig. 1d), and are cleared out within ~40 sec following stimulus end (Fig. 1e). Finally, the setup is also illuminated with an Infrared backlight, to allow optimal segregation of mouse from background and automated analysis of behavior. Mice used in this experiment were 8-16 weeks old males, and all experiments were approved by the Weizmann Institute Animal Care and Use Committee (IACUC).

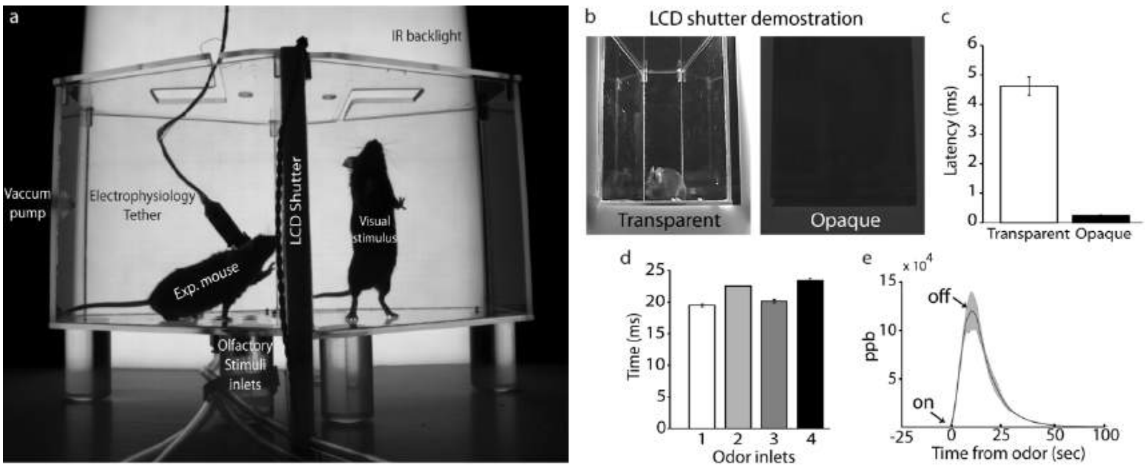


Figure 1. Structure and calibrations of the behavioral apparatus. (a) Major elements of the setup are marked on figure: two transparent cages, one holding the experiment mouse (connected to the electrophysiological recording system), and the other holding the stimulus mouse; an LCD shutter located between cages; four odor inlets; a vacuum pump; an IR backlight illuminating the setup. (b) Demonstrations of the opaque and transparent states of the LCD shutter. (c) Measurements of latency of LCD shutter to shift between opaque/transparent states. (d) Quantification of latency of infused odors to reach

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

experiment cage from each of four odor inlets. (e) PID monitoring of odor concentration following 10 sec induction of ethanol. Values in the figure are presented as mean±SEM.

We utilized this apparatus to record mPFC unit activity in behaving mice introduced with various social and non-social cues, focusing currently on olfactory stimuli. Analysis of unit response patterns revealed that a large fraction of all recorded mPFC units (117/278) responded exclusively to social cues over non-social odors. Many of these units were stimulus-specific, such that 33% responded solely to male cues, while 24% responded exclusively to female odors. We also observed distinct temporal response patterns to male and female cues. Presentation of male odors elicited a robust neuronal response within 150ms of cue delivery, while female-evoked responses developed more slowly, peaking within 1450ms after stimulus onset. High-resolution analysis of behavior revealed that activation of both female-specific and male-specific units, but not non-social units, was significantly correlated with the initiation of approach behavior.

Taken together, our results suggest that the mPFC is involved in encoding salient social cues and in orchestrating appropriate behavioral responses. To the best of our knowledge, this is the first in-vivo evidence for neural representations of social cues in the murine mPFC. Our ongoing experiments focus on exploring social representation in the mPFC of genetic mouse models of autism [8, 9], to uncover how encoding and processing of social cues can guide appropriate and impaired behavioral responses, and contribute to the understanding of the neuronal underpinning of the autism spectrum disorders.

Reference

1. Bicks, L.K., Koike, H., Akbarian, S., and Morishita, H. (2015). Prefrontal cortex and social cognition in mouse and man. *Frontiers in Psychology* **6**.
2. Jenkins, L.M., Andrewes, D.G., Nicholas, C.L., Drummond, K.J., Moffat, B.A., Phal, P., Desmond, P., and Kessels, R.P.C. (2014). Social cognition in patients following surgery to the prefrontal cortex. *Psychiatry Research: Neuroimaging* **224**, 192-203.
3. Wang, F., Zhu, J., Zhu, H., Zhang, Q., Lin, Z., and Hu, H. (2011). Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. *Science* **334**, 693-697.
4. Amodio, D.M., and Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature reviews. Neuroscience* **7**, 268-277.
5. Wong, T.K.W., Fung, P.C.W., Chua, S.E., and McAlonan, G.M. (2008). Abnormal spatiotemporal processing of emotional facial expressions in childhood autism: dipole source analysis of event-related potentials. *European Journal of Neuroscience* **28**, 407-416.
6. Minshew, N.J., and Keller, T.A. (2010). The nature of brain dysfunction in autism: functional brain imaging studies. *Current opinion in neurology* **23**, 124.
7. Zikopoulos, B., and Barbas, H. (2010). Changes in prefrontal axons may disrupt the network in autism. *The Journal of Neuroscience* **30**, 14595-14609.
8. Nakatani, J., Tamada, K., Hatanaka, F., Ise, S., Ohta, H., Inoue, K., Tomonaga, S., Watanabe, Y., Chung, Y.J., Banerjee, R., et al. (2009). Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. *Cell* **137**, 1235-1246.
9. Penagarikano, O., Abrahams, B.S., Herman, E.I., Winden, K.D., Gdalyahu, A., Dong, H., Sonnenblick, L.I., Gruver, R., Almajano, J., Bragin, A., et al. (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235-246.

How to Measure Relations between Personal Values and Behavior Using Experience Sampling Methods

E. Skimina and J. Cieciuch

Institute of Psychology, Cardinal Stefan Wyszyński University in Warsaw, Warsaw, Poland.
e.skimina@uksw.edu.pl

Experience sampling methods (ESM) allow researchers to ask participants about their behavior and feelings in everyday moments of their life [1]. ESM refers to any procedure that has three main qualities: assessment of experience in natural settings, in real time, and on repeated occasions [2]. Assessment in natural settings makes ESM ecologically valid. Because participants report in real time, there is no risk of recall biases. By virtue of its intensive repeated measures design, ESM can be used to study between- and within-person variability [1,2]. ESM can be implemented in a number of ways, including through the use of paper-and-pencil diaries, phone calls, text messages, and smartphone applications [3]. Participants report on their experiences multiple times per day for multiple days [1]. There are three different strategies for creating sampling schedules: signal-contingent sampling, event-contingent sampling, and interval-contingent sampling. When signal-contingent sampling is used, participants receive notifications at random times over the course of the study. This strategy is popular in smartphone-based studies, along with event-contingent sampling. In the second procedure, participants answer some questions following a predefined event. In the third procedure—interval-contingent sampling—participants are asked to answer some questions at regular intervals. This strategy is frequently used in paper-and-pencil studies [1]. A researcher chooses a strategy considering the character of the phenomenon that he or she is interested in. If it is an experience which occurs frequently or is ongoing (e.g., mood), it is better to use signal-contingent sampling. In contrast, when investigated behaviors are rare, the event-contingent sampling strategy is suitable [2].

Our purpose was to adapt experience sampling methodology to the study of relations between personal values and the everyday behaviors that express them. According to Schwartz [4], basic values are trans-situational goals that vary in importance and serve as guiding principles in the life of a person or a group. Personal values are organized into a coherent system that can help to explain individual decision-making and behavior. The values-behavior relations have been studied in retrospective self-report [5,6], but not in real time, which ESM enables. While retrospective self-report gives a global perspective on the relationship between a personal values system and behavior, ESM enables us to follow the dynamic of this relationship by capturing participants' experience as it occurs. To study values-behavior relations in real time, we created the experience sampling form, containing questions about participants' current behavior and the personal values expressed by it.

Because we were interested in measuring behaviors that express personal values, without specifying a priori these behaviors, we asked participants about both: their current behavior and the values expressed by it. We created an ESM form consisting of 10 questions. The first one was "What have you been doing during the past 15 minutes?" and it was open-ended. Then, nine multiple choice questions showed up in random order. All of them began with the text "When you were doing this, how important was it to you to" and ended with a reference to a personal value, for instance "experience something new or exciting," which refers to stimulation value. We referred to 10 values differentiated by Schwartz [4]. The four point scale ranged from *not important at all* to *very important*. We used two strategies—signal-contingent and modified event-contingent—in two studies, each lasting one week. In the first study, participants were prompted at random times in a time window. Using this strategy, we could obtain a representative sample of everyday behaviors. Nevertheless, there was a risk that a participant received notifications while doing something which was not related to any personal value (e.g., sleeping or eating). We attempted to solve this problem in the second study, asking participants to report at least five times a day at various moments chosen by them, while they were engaged in behavior important to them. We expected that they would choose behaviors which they could relate to their personal values. In a typical event-contingent sampling study, participants are asked to report on their experience any time a predefined event occurs (e.g. any time they smoke). We modified this strategy in two ways: (a) by using a broad rather than narrow definition of event—which was “a behavioral act that they considered important and worth the time investment”—and (b) by letting a participant choose only a few of its appearances rather than all of them—we asked our participants to report on five occasions during a day instead of reporting any time they did something important to them.

Both studies were conducted on participants’ own smartphones and created in LifeData software (RealLife Exp mobile application for Android and iOS; <https://www.lifedatacorp.com/>). Approximately 300 people (mostly young adults) participated in the studies.

In our paper we would like to present our conclusions following the two studies and discuss the issues of application of the experience sampling methodology for capturing personal values expression in everyday behavior.

Acknowledgment

This work was supported by Grants 2014/14/M/HS6/00919 from the National Science Centre, Poland.

References

1. Hektner, J.M., Schmidt, J.A., Csikszentmihalyi, M. (2007). *Experience sampling method: measuring the quality of everyday life*. Thousand Oaks, CA: Sage.
2. Conner, T.S., Feldman Barrett, L., Tugade, M.M., Tennen, H. (2007). Idiographic personality: the theory and practice of experience sampling. In R.W. Robins, R.C. Fraley & R.F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 79–96). New York, NY: The Guilford Press.
3. Runyan, J.D., Steinke, E.G. (2015). Virtues, ecological momentary assessment/intervention and smartphone technology. *Frontiers in Psychology* **6**, 1–24.
4. Schwartz, S.H. (1992). Universals in the content and structure of values: theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). New York, NY: Academic Press.
5. Bardi, A., Schwartz, S.H. (2003). Values and behavior: strength and structure of relations. *Personality and Social Psychology Bulletin* **29**, 913–927.
6. Schwartz, S.H., Butenko, T. (2014). Values and behavior: validating the refined value theory in Russia. *European Journal of Social Psychology* **44**, 799–813.

Rodent Behavior Exploration: Discover the Platform

C. Laloux [1], R. Bordet [2], L. Buée [3]

[1]Functional exploration platform, Federative research structure "Neurologic and mental diseases dementia",
[2]JP Aubert research center, INSERM U1172,
[3]Pharmacology department INSERM U1171,
University of Lille North of France, Lille, France – charlotte.laloux@univ-lille2.fr

The Functional Exploration Platform is an area of 150m2 dedicated to the analysis of the behavior of rodents. One part is devoted to the study of the rat and the other for the mice. Platform's behavioral tests allow the exploration of various field of behavior including cognition (attention, different type of memory, executive functions), motor skills and anxiety, to establish the behavioral phenotype associated with rodent models of human neuropathologies.

Global motor activity is evaluated by actimetry in an automated openfield. Coordination, balance and strength are assessed by rotarod, pole test, prehensile traction test and grip test. Anxiety-like behaviour can be assessed by elevated plus maze and light and dark box; and depression-related behaviour by forced swimming test or sucrose preference. Various mazes of increasing difficulty (Y maze, radial arm maze, Barnes maze and Morris water maze) are dedicated to the exploration of spatial memory (working and reference) as well as procedural memory according to the paradigm used. Short term recognition memory can also be evaluated by the object recognition test. Fear conditioning can be used to test passive avoidance or prepulse inhibition behaviour related to emotional, associative learning/memory and attention respectively.

New automated interactive touchscreen technology, analogue to the human CANTAB battery, consists in operant conditioning chambers equipped with a tactile screen. Specific testing paradigmsallow assessment of attention ability through 5 choices serial reaction task, associative learning through object/place paired associated paradigm and mental flexibility through visual discrimination and reversal task (in mice or rats).

Another innovative technology, called Neurologger, allows wireless EEG acquisition during completion of behavioural task/performance of the rodent in behavioral assessment tests, providing a very interesting tool to study cerebral functions.

All experiments protocols were approved by the local Animal Ethical Committee and were carried out in accordance with current French and European Union legislative and regulatory framework on animal experimentation (European Communities Council Directive 86/609). Behavioural test batteries available in the functional exploration platform allow characterization of different models of neuropathologies and the study of pharmacological modulation of their behavioral component. The platform is in permanent evolution and new paradigms can be developed to fit different models depending of behavioral research projects.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

M-kaku: An agent-based model for the emergence of seasonal fission-fusion dynamics in frugivorous primates

R.Dolado[1], E. Gimeno, F.S. Beltran and V. Quera

Institute for Neurosciences, Adaptive Behavior and Interaction Research Group, Department of Behavioral Science Methods, University of Barcelona, Barcelona, Spain. [1]ruth.dolado@ub.edu

Abstract

Social organization, i.e., variations in group size, cohesion and age-sex ratio, is a main factor defining social systems in primates [1]. Specifically, group cohesion reflects group stability over time. Group cohesion can either remain constant with relatively stable composition or groups can exhibit temporal variations in cohesion through fission-fusion dynamics. The term "fission-fusion" was introduced by Kummer [2] to describe social systems observed in some non-human primate species that are characterized by the flexibility of the size of their groups, which change by means of splitting (fission) and merging (fusion) of parties (i.e., subgroups). More recently, [3] proposed a broadening of the term fission-fusion dynamics to reflect the fact that most primates experience temporal fluctuations in group composition. This interpretation of fission-fusion dynamics is expanded to include variation in both spatial cohesion and group composition over time as a population into parties and consequently reconverges as a group [3]. These authors established a framework for categorizing species based on their degree of fission-fusion dynamics. Species and populations are ranked from a low to a high degree of fission-fusion based on three characteristics of the parties: variation in spatial cohesion, variation in party size and variation in age-sex composition.

Different ecological and ethological factors may drive variation in primate group size and could lead to the formation of parties. For example, feeding competition [4-5] and predation pressure [5]. Here, we focus on the management of feeding competition [4-5] as a driving force behind fission-fusion dynamics, and on food availability, which is considered a key factor of feeding competition in primates that affects group size [6], group cohesion [4] and habitat use [7].

The flexible nature of fission-fusion dynamics creates methodological difficulties when defining and measuring group size variation and spatial cohesion in field studies [8]. Moreover, the complexity of foraging environments makes it difficult to measure food availability and distribution [9]. Determining how, why, and which individuals move is often difficult in naturalistic studies. Agent-based models propose a different approach, which makes it possible to predict the variation of fission-fusion dynamics in a group when agents are confronted with a realistic foraging environment.

We have developed *m-kaku* (beta version available on http://www.ub.edu/gcai/group/), an agent-based model implemented in Netlogo (Figure 1). Dyadic social interaction between agents allows the emergence of fission-fusion dynamics and predicts social organization patterns in mangabeys (*Cercocebus torquatus*). The model includes two different elements: food resources and virtual mangabeys. The microworld is a continuous changing landscape formed by a grid of cells which represent food resources in a tropical forest [10]. Each cell contains a level of fruit production defined by the habitat (mangroves, terra fima and coastal palm) and different tree key species that compose the diet of virtual primates. Both habitats and tree species have been chosen to recreate real conditions observed in previous field studies [11]. Virtual mangabeys are represented by black arrows which move around cells looking for food. Social interactions between virtual primates are based on a model proposed by Bonnell [12] for red colobus monkeys (*Procolobus rufomitratus*), and include some new parameters in order to adapt it to the frugivorous behavior observed in mangabeys. Fission-fusion dynamics emerge from individual virtual primates balancing their respective safety (desired neighbors) and energetic requirements (feeding).

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

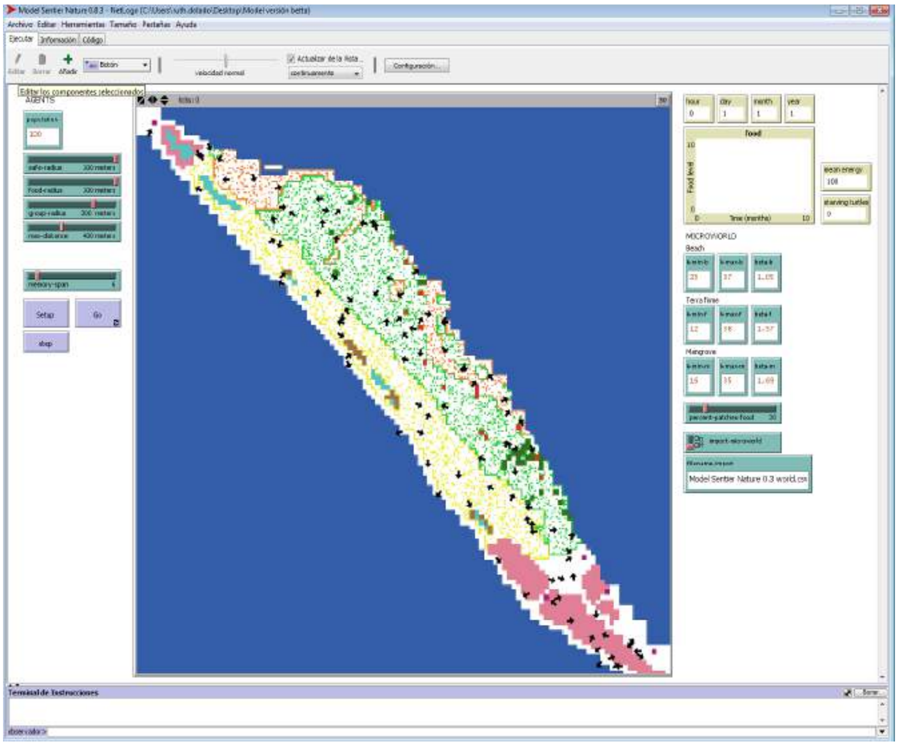


Figure 1. Screenshot of *m-kaku* in Netlogo.

Data on red-capped mangabeys were collected from May to August 2014 in the Sentier Nature forest (2°29' 30″S, 9°44' 56.2″E), close to Loango National Park, Gabon [11]. A single group around 90 individuals was followed during the period of the study to define episodes of fission-fusion. In each encounter with the mangabeys, we estimated party size, GPS locations (during parties movement, we recorded the GPS location every 2 minutes), feeding events occurred on key tree species and, if possible, the party's age-sex composition by direct observation or video recordings. To determine the presence of fruit and its status, several specimens of four species eaten by the resident group (*Sacoglottis gabonensis*, *Guibortia tessmannii, Hyphaene guineensis* and *Manilkara fouilloyara*) were monitored weekly during the period of study.

We assessed the effectiveness of *m-kaku* by comparing simulations results with empirical data obtained from a group of free red-capped mangabeys (*C. torquatus*). We compared different features obtained from both methods as for example, party size proportions, party size depending on food availability, habitat used depending on party size or total home range. Results showed that the *m-kaku* agent-based model reproduces the kind of fission-fusion dynamics observed in the field. Fission-fusion dynamics can be explained by feeding strategies adopted by virtual mangabeys depending on seasonal resource availability. Our software tool can be used to predict social organization patterns in red-caped mangabeys groups, thus contributing to the preservation of endangered communities of this species.

References

1. Kappeler, P.M., van Schaik, C.P. (2002). Evolution of primate social systems. *International Journal of Primatology* **23** (4), 707-739.
2. Kummer, H. (1971). *Primate societies: group techniques of ecological adaptation*. Chicago, Aldine.
3. Aureli, F., Schaffner, C.M., Boesch, C., Bearder, S.K., Call, J., Chapman, C.A., Connor, R., Di Fiore, A., Dunbar, R.I.M., Henzi, S.P., Holekamp, K., Korstjens, H.A., Layton, R., Lee, P., Lehmann, J., Manson, J.H.,

Ramos-Fernandez, G., Strier, K.B., van Schaik, C.P. (2008). Fission-Fusion dynamics. New research framework. *Current Anthropology* **49** (4), 627-654.

4. Janson, C.H., Goldsmith, M. (1995). Predicting group size in primates: foraging costs and predation risks Behavioral. *Ecology* **6**, 326–336.
5. van Schaik, C.P. (1989). The ecology of social relationships amongst female primates In *Comparative socioecology: the behavioral ecology of humans and other mammals* (Standen V, Foley RA, eds.), 195-218. Blackwell, Oxford.
6. Chapman, C.A. (1995). Primate seed dispersal: coevolution and conservation implications. *Evolutionary Anthropology,* **4** (3), 74-82.
7. Dunbar, R.I.M. (1988). *Primate social systems.* Ithaca, Cornell University Press.
8. Chapman, C.A., White, F.J., Wrangham, R. (1993). Defining subgroup size in fission–fusion societies. *Folia Primatologica* **61**, 31–34.
9. Chapman, C.A., Chapman, L.J., Wangham, R., Hunt, K., Gebo, D., Gardner, L. (1992). Estimators of fruit abundance of tropical trees. *Biotropica* **24**, 527–531.
10. Niklas, K. J., Midgley, J.J., Rand, R.H. (2003). Tree size frequency distributions, plant density, age and community disturbance. *Ecology Letters* **6**, 405 411.
11. Dolado, R., Cooke, C., Beltran F.S. (unpublished) How many for lunch today? Seasonal fission-fusion dynamics as a feeding strategy in wild red-capped managbeys (*Cercocebus torquatus*).
12. Bonell, T.R., Campenni, M.C., Chapman, C.A., Gogarten, J.F, Reyna-Hurtado, R., Teichroeb, J.A., Wasserman, M.D., Sengupta, R. (2013). Emergent group level navigation: an agent-based evaluation of movement patterns in folivorous primate. *PLOS ONE,* **8** (10).

A Novel Automatic and Affordable System for Tracking Social Preference in Rodents

S. Netser, S. Haskal, M. Levy and S. Wagner

Sagol Department of Neurobiology, University of Haifa, Haifa, Israel. shainetser@gmail.com

Social neuroscience, one of the most rapidly developing fields in neuroscience, commonly uses social behavior of rodents as an animal model.. Across the years, many laboratory assays for social behavior in rodents were developed, including the social preference, avoidance and recognition tests. In the past, most studies used human observers for evaluation and quantification of rodents' social behavior [4]. This methodology is inaccurate due to the dependency on the subjective estimation of the observer, and in many cases can lead to inconsistent results. Another disadvantage of this methodology is the difficulty to reliably evaluate the dynamics of the behavior. Several outomated systems were previously presented and used for tracking mice and rats social behavior, but those systems are either unaffordable for all labs or require complex custom made setups [1,2,5].

Here we aimed to automatize the analysis of such behaviors using a new experimental setup and a simple and affordable custom-made tracking system. All the experiments were carried on C57B6 mice and SD rats and were approved by the institutional Animal Care and Use Committee of the University of Haifa. The experimental setup consists of a white Plexiglas box placed at the middle of an acoustic chamber which is equipped with a high-quality monochromatic camera (Point Grey) at the top (Figure 1a). Inside the box we placed two Plexiglas triangular chambers in two randomly selected corners, into which stimuli mice or objects can be placed. Two versions of the chambers were used. In the first one, three slits were grooved towards the center of the box, thus enabling a subject rodent to investigate the stimulus rodent in a limited way through the slits. The second version was similar but used a metal net placed instead of the slited area, thus enabling a more direct interaction between the animals via the net. The camera is equipped with a wide angle lens which enables a view of the subject rodent investigation directed towards the stimulus in the chamber. The advantages of using this arrangement are first the ability to view in an optimal way the social interaction area which simplifies the determination of investigation behavior. Second, it enables to randomly place the chambers in different corners of the box, thus minimizing the role of spatial navigation and place preference during repeated trials. Finally, the setup is suitable for easily placing an anesthetized stimulus against the inner wall of the chamber, thus neutralizing the stimulus behavior while allowing investigation of its body by the subject.

In addition to the setup, custom-made software was written in MatLab for tracking the rodent subject location and it's interactions with the stimuli located inside the chambers. After uploading a movie file, the experimenter defines areas to be considered as "compartments" or "stimuli" (Figure 1b). For "Compartments", the software tracks the time points in which the subject is found inside each one of them, while "Stimuli" are areas which the software tracks contact with (Figure 1c). Two algorithms can be selected for analysis of a given experiment: tracking a freely-moving rodent and tracking a rodent connected to a cable. Although the algorithms are simple, in the suggested setup we found a goodcorrelation between the investigation times measured by observer and by the softwares (88%). False alarms, resulting from random touches of the subject with the chambers were relativly rare and thus the difference between the software's and observer's measurments were minor. Moreover, we are now testing a new algorithm which detects the rodents' head and identify its contact with the chambers. Preliminary results show similar detection capabilities for mice as the previous algorythm. In the software interface, the parameters for analysis can be defined and uploaded for a batch of movie files and the results can be later on presented with several quantifications possibilities. Finally, the results files from several animals can be collapsed for further analysis in both MatLab and Excel.

Using this system we could reliably analyze several well-known paradigms, such as the social preference and social recognition tests, in both mice and rats [4]. A precise evaluation of the behaviors across time, using this system, showed a specific and consistent exploration dynamics for both paradigms. To validate the results we used this system to evaluate the effect of blocking the activity of oxytocin [3], a hypothalamic neuropeptide which is well-known to regulate social behavior and memory, by i.p. administration of the oxytocin receptor antagonist L-368,899 to mice. The automatic system could detect the same reduction in sociability due to the blockade of oxytocin activity as was previously shown by observers, and to evaluate its effect on the behavior dynamics. We also used the setup in rats to show their social preference for an anesthetized stimulus over an object. In future experiments we aim to study the preference between different anestetized stimuli (male/female, juvenile/adult) and expaned the analysis to their body parts (head, thorax, anogenital). Finally, we show that the system is capable of tracking social behavior while the rodents are connected to an optic fiber or a cable-based electrophysiological recording system.

Figure 1:

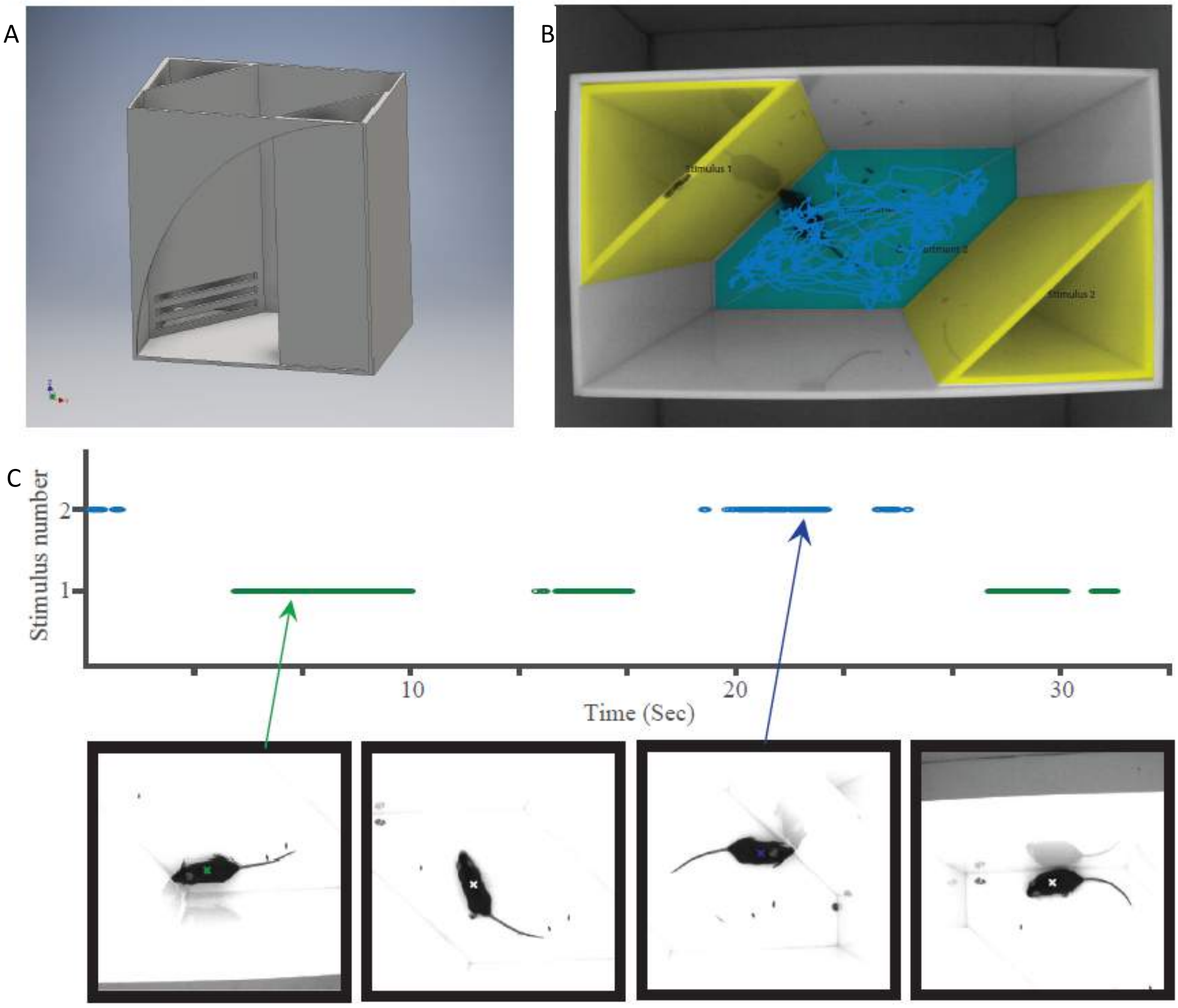


Figure 1. Social behavior setup and tracking algorithm. A. Illustration of the setup box (37 X 22 X 35cm) and the triangulated chambers with slits (12cm isosceles, 35cm hight). B. An example of one of the software outputs. The mouse trace is plotted on the background of the arena and the defined areas for analysis are marked. C. An example result of one of the algorithms tracking the rodent interaction times with the stimuli inside the chambers.

Reference

1. Hong W, Kennedy A, Burgos-Artizzu X.P, Zelikowsky M, Navonne S.G, Perona P, Anderson D. J (2015). Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proc Natl Acad Sci U S A*. 22: 112(38): E5351-60.
2. Nadler J.J, Moy S.S, Dold G, Trang D. Simmons N, Perez A, Young N.B, Barbaro R.P, Piven J, Magnuson T.R and Crawley J.N (2004) Automated apparatus for quantitation of social approach behaviors in mice. *Genes Brain Behav*. 3: 303-314.
3. Quattropani A, Dorbais J, Covini D, Pittet PA, Colovray V, Thomas RJ, Coxhead R, Halazy S, Scheer A, Missotten M, Ayala G, Bradshaw C, De Raemy-Schenk AM, Nichols A, Cirillo R, Tos EG, Giachetti C, Golzio L, Marinelli P, Church DJ, Barberis C, Chollet A, Schwarz MK (2005). Discovery and development of a new class of potent, selective, orally active oxytocin receptor antagonist. *J Med Chem.* 48(24):7882-905.
4. **Moy S. S,** Nadler J.J, Perez A, Barbaro R.P, Johns J.M, Magnuson T.R, Piven J. and Crawley J. N. (2004). Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. *Genes Brain Behav.* 3, 287–302.
5. Noldus L.P, Spink A.J and Teqelebosch R.A (2001). Ethovision: a versatile video tracking system for automation of behavioral experiments. *Behav Res Methods Instrum Comput*. 33(3): 398-414.

ART – Automated Rodent Tracker

B.M. Hewitt[1,2], MH. Yap[2] and R.A. Grant[1]

[1]School of Science and the Environment, Manchester Metropolitan University, Manchester, UK, Brett.Hewitt@mmu.ac.uk; Robyn.Grant@mmu.ac.uk
[2]School of Computing, Mathematics, and Digital Technology, Manchester Metropolitan University, Manchester, UK, M.Yap@mmu.ac.uk

Abstract

Locomotion and behavioral analysis in rodents is becoming an increasing popular research area [1], particularly with the shift towards non-invasive monitoring of behavior to assess rodent health and welfare. Popular behavioral tests including open field and novel object exploration methods. These tests usually involve the observers manually annotating movements and behavioral interactions, using programs such as Noldus Observer[2], which only gives semi-automated information, such as duration and frequencies of certain behaviors, including object contacts and wall-following. With the advent of new and freely available computer vision libraries [3, 4], many software packages have been developed to automate this process [5, 6, 7]; however, the wealth of behavioral data remains semi-quantitative in essence. The automatic software packages have other limitations too. They often cannot be applied more generally to experimental set-ups (over different lighting schedules, with different objects present etc) and also require the user to manually enter multiple settings. This means that the programs might work successfully for a handful of datasets or clips, but fail on others (Fig 1).

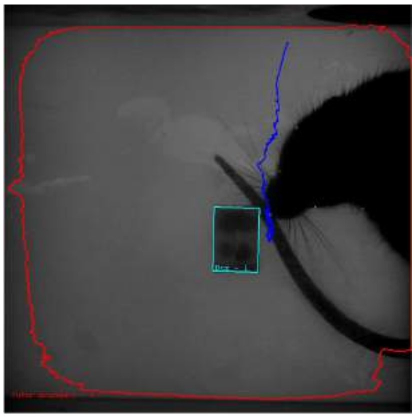


Fig 1 – ART annotation of a difficult video, the dark blue line represents the motion track.

Here, we present the ART (Automated Rodent Tracker), which has been developed to overcome the limitations of existing automatic behavioral software. ART boasts an intuitive user interface, requiring little user input. Feedback from existing programs have been that altering settings is not intuitive for users, therefore, ART produces visual feedback to allow the user to enter the correct value without any prior knowledge of what the setting actually does. Using a modified background subtraction technique (motion detection [8] to get movement, minimum convex hull [9] to get area of motion, novel method to bend hull to preserve static objects that lie within the convex area), the outside boundaries and objects of interest are automatically identified. Finally, another novel technique finds the location of the rodent's nose-tip quickly, automatically and to a high level of accuracy. Using information from the rodent and object positioning, the program will also automatically provide information about the time and duration of object and wall interaction behaviors. In clips where the body and tail are clearly visible, an automated detection process uses the rodent body contour to perform skeletonisation [10] to identify the mid-line along the body. Using the distance from the mid line to the mouse contour, sections of the head, body (waist and hips), and tail can all be identified, in order to estimate body size, or pelvic area volume, for instance. Easy export options are provided for the raw pixel data, locomotion

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

movement analysis, rodent size approximations and also the interaction behaviors. The software has been extensively tested and provides accurate tracking even in some of the most difficult datasets, which include videos where the rodent interacts with objects, therefore making the rodent and objet appear as one entity or videos where objects partially obscure the rodent. For example, Fig 1 shows the nose partially obscured by the tail and Perspex block in some portions of the clip.

This program provides a freely-available alternative to other automatic behavioral programs. It is easy to use for non-specialists and exports in to general programs, such as excel. The categorization of behaviors provides the ability to quickly and easily process the results from video footage of open field and novel object tasks, by providing frequency and duration information about exploratory behaviors and positioning. It also provides quantification of locomotion parameters, such as displacement, orientation and speed. This program functions quickly, easily and robustly over a range of challenging videos, and provides both quantitative and semi-quantitative measurements of rodent behavior for locomotion and exploration.

References

1. Grant, R.A., Sharp, P.S., Kennerley, A.J., Berwick, J., Grierson, A., Ramesh, T., & Prescott, T.J. (2014) Abnormalities in whisking behaviour are associated with lesions in brain stem nuclei in a mouse model of amytrophic lateral sclerosis. *Behavioural brain research,* **259***,* 274-283.
2. Noldus, L.P. (1991) The Observer: a software system for collection and analysis of observational data. *Behavior Research Methods, Instruments and Computers,* **23(3),** 415-429.
3. Chang, P.L., & Teng, W.G. (2007) Exploiting the self-organizing map for medical image segmentation. *Computer Based Medical Systems 2007. Twentieth IEEE International Symposium,* 281-288.
4. Sherwood, T., Ahmad, E., & Yap, M.H. (2015) Formulating efficient software solution for digital image processing system. *Software: Practice and Experience.*
5. Perkon, I., Košir, A., Itskov, P.M., P.M., Tasič, J., & Diamon, M.E. (2011) Unsupervised quantification of whisking and head movement in freely moving rodents. *Journal of neurophysiology* **105(4),** 1950-1962
6. Clack, N., O'Connor, D., Huber, D., Petreanu, L., Hires, A., Peron, S., … & Myers, E. (2012) Automated tracking of whiskers in videos of head fixed rodents. *PLOS Computational Biology,* **8(7),** e1002591
7. Knutsen, P.M., Derdikman, D., & Ahissar, E. (2005) Tracking whisker and head movements in unrestrained behaving rodents. *Journal of neurophysiology,* **93(4),** 2294-2301
8. Illingworth, J., & Kittler, J. (1998) A survey of the Hough transform. *Computer vision, graphics, and image processing,* **44(1),** 87-116
9. Getz, W.M., & Wilmers, C.C. (2004) A local nearest-neighbour convex-hull construction of home ranges and utilization distributions. *Ecography,* **27(4)**, 489-505
10. Zhang, T. Y., & Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, ***27*(3)**, 236-239.

We Can See Comfort Now…
How Measuring Behaviour Might Help to Objectively Assess Comfort Perception

J.M.A. Bouwens[1,2], U.W. Schultheis[1], P. Vink[2]

[1]Zodiac Seats US, Department of Human Factors and Ergonomics, 2000 Weber Dr, Gainesville, TX 76240, USA, Joyce.Bouwens@zodiacaerospace.com
[2]Delft University of Technology, Faculty of Industrial Design Engineering, Landbergstraat 15, 2628 CE Delft, The Netherlands

Interior designers might choose a chair mainly because of its aesthetics. However, comfort seems to be a more important factor for the user of the seat, especially when they have to sit for a long period of time, for example during a long-haul flight. Relaxation and psychological well-being are key aspects of comfort experience [1] [2] [3], however aesthetics can influence the perceived comfort (e.g. a 'nicely' coloured seat is perceived as more comfortable than an 'ugly' coloured seat [4]). Next to the aesthetics, other subjective factors, such as proxemics, pleasure and association [5] might affect comfort experience. Therefore other, more objective, methods are welcome to assess comfort.

This case study researched the comfort provided by travel pillows used in airplane seats. Six different travel pillows were used by 10 participants. While using each pillow for 45 minutes, the head movements of the participants were video recorded. After using all the pillows, participants ranked the travel pillows (e.g. which pillow was the most comfortable one and which pillow was the worst?) and the researcher coded the head movements (direction and duration) with The Observer XT 12.5 (Noldus Information Technology).

The correlation between the number of head movements and the comfort ranking is analysed with a Spearman rank-order test in SPSS 22 (IBM). A moderate (correlation coefficient between 0.5 and 0.7) and strong (correlation coefficient between 0.7 and 1.0) uphill positive relationship between comfort ranking and head movements are measured for respectively 3 and 4 out of 10 participants (see Figure 1).

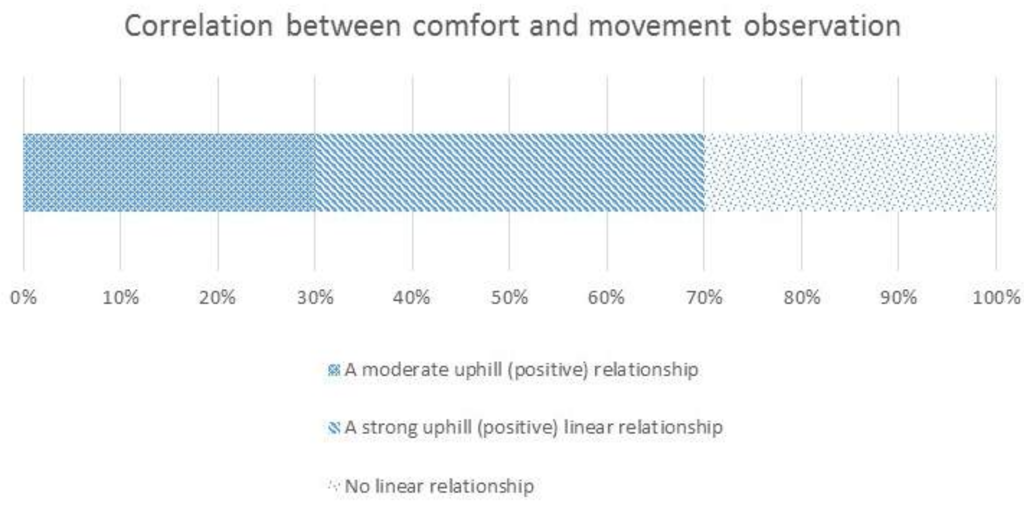


Figure 1. Correlation between comfort and movement observation.

Although the number of participants was limited and not all influencing external factors could be precluded, this study does indicate a correlation between the amount of movement and comfort experience. In this experiment the relationship is showed for 70% of the cases, which might indicate that observing the head and neck area might not be representative enough to assess comfort. However, previous findings [6] in where a relation between in chair movement of the whole body (measured by force sensitive resistors) and comfort experience

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

was established. The results of this explorative study therefore support that behavioural research could contribute to an objective measuring method for the assessment of comfort.

References

1. Helander, M. and Zhang, L. (1997). Field studies of comfort and discomfort in sitting. *Ergonomics* 40(9), 895-915.
2. De Looze, M.P., Kuijt-Evers, L.F.M., Van Dieën, J. (2010), Sitting comfort and discomfort and the relationships with objective measures. *Ergonomics* 46(10), 985-997.
3. Vink, P., Brauer, K. (2011) Aircraft interior comfort and design. *Taylor and Francis group*.
4. Bronkhorst, R. et al. (2001) Emotion and comfort in cabins. *TNO, Hoofddorp*, Publ. nr. R2014871/4020054.
5. Ahmadpour, N. et al. (2014) The thematic structure of passenger comfort experience and its relationship to the context features in the aircraft cabin. *Ergonomics* 57(6), 801-815.
6. Cascioli, V. et al. (2016) A methodology using in-chair movements as an objective measure of discomfort for the purpose of statistically distinguishing between similar seat surfaces. *Applied Ergonomics* 54, 100-109.

Studying Episodic Memory of Rats in the Morris Water-maze

A.J. Ernyei, E. Császár, I. Plangár, F. Kassai, I. Gyertyán

MTA-SE NAP B Cognitive Translational Behavioural Pharmacology Group
Department of Pharmacology and Pharmacotherapy, Semmelweis University
Institute of Cognitive Neuroscience and Psychology, Research Center for Natural Sciences, MTA
Budapest, Hungary. ernyei.aliz@med.semmelweis-univ.hu

Introduction

Episodic memory is the ability to mentally relive one's own past events. Brain damage or dementias such as Alzheimer´s disease frequently lead to the impairment of episodic remembering therefore several tests were developed to study this cognitive skill in humans.

Aim

Our aim was to translate this test to rats and query them whether they can determine which time of the day they are. For this purpose we modified the Morris water navigation task in that the platform was placed onto different locations depending on the time of the day.

Methods

12 male 3 months old Lister Hooded rats were used which were kept in reversed light-dark cycle (dark active phase from 4:00 am until 4:00 pm) and were fed restricted at 3:30pm, at the end of their day.

Standard test

First, these animals were trained in four days (3 sessions daily) to escape onto a hidden platform from any starting position in 190 cm diameter Morris water maze. The platform with 10 cm diameter surface was placed in the south-east quadrant of the pool.

Experiment 1

Then we modified the standard test: the rats were swimming three sessions per day: at 5:00 am ('morning' for rats), 10:00 am ('noon') and 3:00 pm ('evening'). The platform was located in north-east in the maze every morning, north-west at noon and south-west in the evening 40 cm away from the outside edge of the pool. Each session consisted of two trials: the first one was the 'query trial'; performance of the animals in this trial formed the outcome measures of the study. The second trial served as a reinforcing trial regarding the daytime position of the platform. The animal was placed at a random cardinal starting position (north, west, south or east) at the beginning of each trial. Escape latency to the platform and time of first visits to the other possible target zones were measured and pathway of swimming was tracked. The rats were trained for 2x4 days with a weekend in between. The primary outcome measure was the rank of visit to the current target zone among the first visits to the other possible target locations.

Experiment 2

In order to facilitate the performance of the animals we simplified the task. We reduced the sessions per day to a noon session at 9:00 am and an evening session at 3:00 pm. The platform remained in the north-west quadrant in the noon and in the south-west quadrant in the evening session and the rat was put into the pool each time at the east starting point. In case of swimming directly to the platform, without entering the other possible target zone, they got three pellets as a reward. The test was repeated 2x4 days long.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Results

Standard test
All the rats successfully completed the standard Morris water-maze test with an average escape latency of 12.1± 2.79 sec on the last day.

Experiment1
The animals learned in the first week that there were only three options to escape and from the second week theyignored the fourth, formerly used south-east platform position. They followed the strategy of sequentially visiting the three potential target zones to find the platform, which took them about 7 seconds, while swimming directly to the target was possible in about 4 seconds. Therefore, we only calculated with three possible locations (the ones used in the morning, noon and evening) in determining the rank order of visits to the possible target zones. Theoretically, if rats randomly select among the three possible target zones this value is 2 on average, while they perfectly know the actual location it is 1 (i.e. they directly swim to the platform). Results showed, however, that the animals were not able to find out the contextual rule of the platform location; they found the actual target roughly by the 2nd choice (the mean rank was 1.8± 0.07) in week 2. However, more detailed analysis revealed that performance of the animals were significantly better than chance in the morning and at noon but not in the evening (the mean ranks are 1.7± 0.12; 1.6± 0.10; 2.2± 0.12, respectively).

Experiment 2
Again, after a few days rats adapted to the two possible platform sites and ignored the previous morning location. They have found the platform in 6 seconds when they sequentially visited both possible target zones. The calculated scores were between 1 and 2 as there were 2 possible platform locations in this experiment. The score of noon sessions (1.3± 0.07) again significantly differed from chance level which was 1.5 in this setup. However, despite the longer time interval between the daily sessions and the radically simplified task, they still did not manage to discover the rule in the evening session (1.6± 0.09, ns. from chance). Repeating the experiment after one month break the data were essentially similar (the mean ranks are 1.3± 0.06, $p<0.05$ and 1.5± 0.07, ns. at noon and in the evening; respectively).

Conclusion

The rats were able to learn in a few days on which locations to search for the platform. As the possible target zones are positioned about 1-2 sec swimming distance from each other for the rat, probably it is a smaller mental effort to follow the strategy of sequentially visiting the potential target zones than searching for the rule to find the target directly.

We assume that by increasing the motivation to successfully complete the episodic memory task may improve the performance of the animals. A possible way for doing that is to augment the cost of correcting an initial wrong choice, e.g. increasing swimming distance between the possible target zones. We intend to expand our experimentation into this direction.

Ethical statement

The experiments were authorized by the regional animal health authority in Hungary (resolution number PEI/001/3572-4/2014) and conformed to the Hungarian welfare law and the EU 63/2010 Directive.

Acknowledgement

This study was funded by the Hungarian National Brain Research Program (NAP), contract # KTIA_NAP_13-2014-0015.

Assessing Oculomotor Function in Rett Syndrome Using Integrated EEG and Eye Tracking Technology

B.J. Byiers[1], F.J Symons[1], C.A. Nelson[2], and J.T. Elison[3]

[1]Department of Educational Psychology, University of Minnesota, Minneapolis, USA.
byier001@umn.edu
[2]Boston Children's Hospital and Harvard Medical School, Boston, USA.
[3]Institute of Child Development, University of Minnesota, Minneapolis, USA.

Eye movement paradigms are being increasingly used to evaluate the functional integrity of brain systems involved in sensorimotor and cognitive processing in a range of clinical disorders [1]. Modern eye tracking technology produces reliable, sensitive, non-invasive measurement, and many tasks are appropriate for challenging populations, such as infants and individuals with disabilities, as they do not require conscious motor control and can be designed to limit cognitive demands. The neurophysiologic and neurochemical bases of oculomotor control have been well characterized, making it possible to relate performance to specific brain regions subserving discrete aspects of eye movement control [2].

Oculomotor function has been shown to be a highly sensitive and reliable indicator of cognitive function in a range of clinical disorders, including schizophrenia, Parkinson's disease, and traumatic brain injury [3-5]. Studies in such clinical disorders, as well as single electrode recording and lesion studies in primates, have resulted in strong evidence of differentiated pathways associated with different types of oculomotor tasks. For example, saccadic hypometria, or undershooting a visual target during eye tracking tasks, is a hallmark of disorders of the cerebellum [6], whereas lesions to the parietal eye fields result in significant increases in latencies of visually triggered saccades.

Rett syndrome (RTT) is an ideal disorder in which to test the promise of eye movement paradigms as measures of neurological function. RTT is a devastating neurodevelopmental disorder that leaves affected girls and women unable to speak or use their hands. Identification of mutations of the X-linked methyl-CPG binding protein (MeCP2) gene as the cause of most cases of RTT has resulted in promising advances toward identifying therapeutics for treating core symptoms of RTT [7]. Evidence from human and animal model studies suggest that deficits in MeCP2 protein levels likely result in intellectual impairment [8]. Separating the effects of cognitive function alone from those of motor impairment, however, is nearly impossible using traditional assessment methods. Following the regression period, most individuals with RTT obtain scores equivalent to 6- to 18-month old developmental levels on standardized measures [9]. Some researchers have argued that these assessments may underestimate cognitive capacity in this population, and that adaptations to assessments, such as longer response times and acceptance of idiosyncratic responses such as 'eye gaze referencing' may result in higher scores [10]. Recently, researchers have suggested that that eye tracking may be a feasible way to bypass the impaired motor system, providing a more reliable method of evaluating cognitive functions in this population [11].

Several researchers are currently conducting studies using eye tracking to evaluate cognitive functions, such as attention and memory, in RTT. A potentially significant limitation to these studies, however, is that the status of oculomotor function in RTT is unknown. Anatomical studies show that MeCP2 selectively affects specific neural structures/regions that could plausibly lead to deficits in oculomotor control [12, 13], and clinical observations support the hypothesis of altered oculomotor control [14, 15]. We hypothesize that oculomotor performance could be a sensitive and biologically relevant marker of cortical function in this population. This project is designed to test specific hypotheses regarding the integrity of the oculomotor system in girls and women with Rett syndrome (RTT). The overall goal of the project is to use an integrated system of electroencephalography (EEG) and eye tracking to assess the behavioral and cortical systems implicated in stimulus-driven, reflexive eye movements.

Methods

A sample of females with classic Rett syndrome (of all ages) and a sample of age-matched typically-developing females are currently being recruited for the study. All methods were approved by the University's Internal Review Board.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Equipment. All assessments are administered on a Tobii TX300 eye tracker integrated with a high-density, 128 channel EEG system (EGI Inc. with NetStation software).This system has been used extensively with infants, and with developmentally disabled populations due to its quick application and the lack of need for skin abrasion or other intrusive procedures. Experimental procedures and data collection from both systems are controlled through MatLab, using Psychtoolbox-3 extensions. During the assessments, participants will be seated in a supportive chair or their own assistive wheelchair chair, approximately 65 cm from the eye tracker.

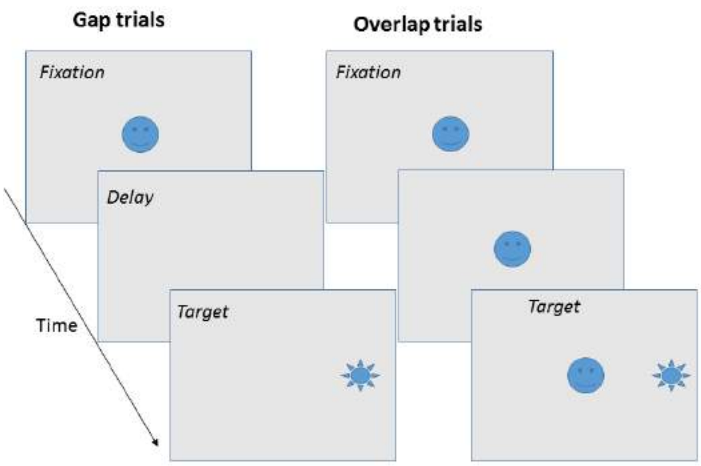


Figure 1. Example presentation sequences for gap and overlap trials. The central stimulus (face) disappears prior to the appearance of the peripheral target (star) on gap trials (left side), but remains for the duration of the overlap trials (right side).

Gap-overlap task. This well-established, eye-tracking task will be used as a behavioral measure of oculomotor function. The tasks consists of two types of trials in which an initial, central stimulus is presented, followed by presentation of a second stimulus on the periphery of the screen, either concurrent with the central (overlap trial), or after the central stimulus has disappeared for a brief period (gap trial) (see Figure 1). Numerous studies have shown that latency to initiate saccadic eye movements are significantly reduced in gap trials versus overlap trials. Oculomotor efficiency, as measured by average latency of saccadic eye movements during the gap trials, results from temporary inhibition of fixation neurons in the rostral pole of the superior colliculus and disinhibition of the saccade-generating circuit, mediated in part by the reticular formation in the brainstem [16]. In 7-month-olds, oculomotor latencies in the gap condition are uniquely associated with the microstructural properties of white matter ascending dorsally from the brainstem [17]. In contrast, saccadic latencies in the overlap condition reflect a combination of oculomotor responding and attentional orienting [16], and are uniquely associated with white matter microstructure in the splenium of the corpus callosum [17]. The link between orienting and the splenium has also been corroborated in adults [18]. Therefore, differences in latencies between the gap and overlap trials will provide important information regarding the specific circuitry and processes that are affected. Saccadic latencies in the gap and overlap trials, as well as the average difference between the two conditions will serve as the dependent variables for this task.

EEG measures. Visual acuity and processing will be assessed using a checkerboard reversal pattern to elicit visual evoked potentials (VEP). Stimuli will be presented with 100 ms reversals, and EEG will be sampled at 500 Hz. VEP will be recorded over the occipital scalp, with peak amplitude and latency as the dependent measures. To minimize loss of trials due to inattention, the experiment is set-up to automatically present VEP trials when the eye tracker records that the participant's eyes are fixated on the screen. Event-related potentials (ERPs) will also be derived from the gap-overlap stimulus presentations in order to evaluate the presence and form of the pre-saccadic spike potential, an event-related potential that is believed to reflect cortical coordination of saccadic eye movements and is seen under most saccade situations in typically developing individuals beginning at 12 months of age [20].

Expected outcomes

Data are currently being collected for the study. We hypothesize that these procedures will reveal anomalies in the oculomotor system among individuals with RTT. Specifically, we expect that (1) the gap-overlap task will show increased saccadic latencies during gap trials compared to typically developing individuals, (2) EEG measures will reveal the presence of a pre-saccadic spike potential among individuals with RTT, but the amplitudes are expected to differ from those of the comparison group, and (3) saccadic latencies during gap trials will be significantly correlated with the amplitude of the pre-saccadic spike potentials. These measures may also be significantly correlated with the amplitude and latencies of visual evoked potentials, which would indicate that observed deficits in oculomotor function result from cortical dysfunction rather than simple ocular muscle weakness.

References

1. Reilly, J.L., Lencer, R., Bishop, J.R., Keedy, S., & Sweeney, J. A. (2008). Pharmacological treatment effects on eye movement control. *Brain and cognition*, *68*(3), 415-435.
2. Leigh, R.J., & Kennard, C. (2004). Using saccades as a research tool in the clinical neurosciences. *Brain, 127,* 460-477
3. Buhmann, C., Kraft, S., Hinkelmann, K., Krause, S., Gerloff, C., & Zangemeister, W. H. (2015). Visual Attention and Saccadic Oculomotor Control in Parkinson's Disease.. *European neurology, 73,* 283-293.
4. Hutton S., & Kennard, C. (1998). Oculomotor abnormalities in schizophrenia: A critical review. *Neurology, 50,* 604-609;
5. Kraus M. F., Little, D. M., Donnell, A. J., Reilly, J. L., Simonian, N., & Sweeney, J. A. (2007). Oculomotor function in chronic traumatic brain injury. *Cognitive and Behavioral Neurology, 20,* 170-178.
6. Robinson F. R., & Fuchs, A. F. (2001). The role of the cerebellum in voluntary eye movements. *Annual review of neuroscience, 24,* 981-1004.
7. Amir R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., & Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature genetics*, *23*, 185-188.
8. Santos M., Summavielle, T., Teixeira-Castro, A., Silva-Fernandes, A., Duarte-Silva, S., Marques, F. Martins, L., et al. (2010). Monoamine deficits in the brain of methyl-CpG binding protein 2 null mice suggest the involvement of the cerebral cortex in early stages of Rett syndrome. *Neuroscience, 180,* 453-467.
9. Byiers B.J., & Symons, F.J. (2012). Issues in Estimating Developmental Level and Cognitive Function in Rett Syndrome. *International Review of Research in Developmental Disabilities, 43,* 147-185.
10. Demeter K. (2000). Assessing the developmental level in Rett syndrome: an alternative approach? European Child & Adolescent Psychiatry, 9(3), 227-233.
11. Rose S.A., Djukic, A., Jankowski, J.J., Feldman, J.F., Fishman, I., & Valicenti-McDermott, M. (2013). Rett syndrome: An eye-tracking study of attention and recognition memory. Developmental Medicine and Child Neurology, 55, 364-371.
12. Reiss A. L., Faruque, F., Naidu, S., Abrams, M., Beaty, T., Bryan, R.N., & Moser, H. (1993). Neuroanatomy of Rett syndrome: a volumetric imaging study. Annals of neurology, 34(2), 227-234.
13. Takakusaki K., Saitoh, K., Harada, H., & Kashiwayanagi, M. (2004). Role of basal ganglia-brainstem pathways in the control of motor behaviors. *Neuroscience Research, 5,* 137-151.
14. Einspieler, C. Kerr, A.M., & Prechtl, H.F. (2005). Abnormal general movements in girls with Rett disorder: the first four months of life. *Brain and Development*, *27*, S8-S13 et al., 2005;
15. Neul J.L., Kaufmann, W.E., Glaze, D.G., Christodoulou, J., Clarke, A.J., Bahi-Buisson, N., et al. (2010). Rett syndrome: Revised diagnostic criteria and nomenclature. *Annals of Neurology, 68*, 944-950
16. Dorris, M.C., & Munoz D.P. (1995). A neural correlate for the gap effect on saccadic reaction times in monkey. *Journal of Neurophysiology, 73,* 2558–2562.
17. Elison J.T., Paterson, S.J., Wolff, J.J., Reznick, J.S., Sasson, N.J., Gu, H., et al. (2013). White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. *American Journal of Psychiatry, 170,* 899-908
18. Niogi S., Mukherjee, P., Ghajar, J., & McCandless, B.D. (2010). Individual differences in distinct components of attention are linked to anatomical variations in distinct white matter tracts. *Frontiers of Neuroanatomy, 4*, 2
19. Csibra, G., Tucker, L.A., Volein, A., & Johnson, M.H. (2000). Cortical development and saccade planning: The ontogeny of the spike potential. *Cognitive Neuroscience, 11,* 1069-1073.

Welfare assessment in Wild Species: New non invasive Methods to Measure Stress in African Grey Parrots (Psittacus Erithacus)

M. Alnot-Perronin, C. Bienboire-Frosini, C. Chabaud, A. Cozzi and P. Pageat

IRSEA Research Institute in Semiochemistry and Applied Ethology, Quartier Salignan, 84400 Apt, France
m.alnot@group-irsea.com

The authors declare that there is no conflict of interest.

Because the number of pet parrots is increasing, and because their sensitive and intelligent nature makes them particularly vulnerable to stress, it is important to assess their welfare using the least invasive technics.

In a previous work on a population of African grey parrots (Psittacus erithacus) we used the heterophil to lymphocyte ratio (HLR) to assess stress as it is used in poultry or free-living birds. One drop of blood was necessary from each individual. HLR measures were obtained analysing blood smears: the count and ratio of these blood cells was performed to compare two groups of parrots: a group showing anxiety signs and a group showing no anxiety signs. The HLR was found different between the groups, significantly higher in the group showing stress signs (*Mann-Whitney U test:* $U=71.5$; $p=0.015$) and the correlation between HLR and the number of anxiety signs was significant (*Spearman's rho:* $\rho=0.72$; $p<0.05$) showing the relevance using this measure of stress in our population. But performing blood veinipunctures even taking one only drop could be stressful for some individuals because it demands to restrain them, inducing a risk of injuries for the parrot or even for the investigator. Therefore, a less invasive measurement technic was preferable. As the use of glucocorticoid metabolites (GCM) from mammalian feces and bird droppings is now widely used to assess stress and welfare issues in various species, especially in wild or endangered ones, we decided to use it in our population of African Grey parrots. Several authors have highlighted the requirement of a rigorous validation of GCM measurement for each new studied species. But to our knowledge, no such work has yet been published for African grey parrots despite the high interest to monitor their welfare and stress-related issues. Procedure validation includes among others the selection of a suitable combination of GCM extraction and immunoassay methods. During a test where the parrots where not put in a new cage, we collected the droppings of our population. Our team evaluated several method combinations to extract and assay GCM from a pool of 18 droppings from both sexes of African grey parrots: (i) two pre-extraction treatments of droppings (fresh vs. dry); (ii) two extraction buffers (60% ethanol vs. 60% methanol); (iii) three commercially available enzyme immunoassay (EIA) kits initially designed for assaying corticosterone (from Enzo Life Sciences, Cayman Chemical, and Immuno Diagnostic Systems). To select and analytically validate the best global procedure, several quality parameters were assessed. The standard analytical validation showed that the best procedure to measure GCM in African grey parrots droppings is a combination of an extraction with 60% methanol and a measurement using the Corticosterone EIA kit from Cayman Chemical Company from fresh or dry droppings. Biological validation was shown through the establishment of a significant correlation (Pearson coefficient correlation = 0.48; p = 0.0082) between HLR and GCM on a dataset from 29 individuals (14F; 15M).

This study highlights the importance of species specific validated non invasive measures in the evaluation of parrots' stress in order to improve the clinical approach for these species. These methods could eventually allow to estimate the kind of welfare we offer to psittacine birds when we are keeping them as pets.

Monitoring Parkinson's Disease Progression Using Behavioural Inferences and Smartphones

Julio Vega, Markel Vigo, Caroline Jay and Simon Harper

School of Computer Science, University of Manchester, Manchester, UK ({julio.vega, markel.vigo, caroline.jay, simon.harper}@manchester.ac.uk)

This research aims to monitor Parkinson's Disease (PD) in a longitudinal, naturalistic, non-disruptive and non-intrusive way. It uses smartphones to log social, environmental, and interaction data about patients and their surroundings. This data is then processed to infer a set of behavioural metrics (a latent behavioural variable or LBV) of people's activities and habits. Then, the LBV's trends are quantified and mapped to the progression of the disease. As a part of the pilot study to test the proposed methodology, we collected ~290 million records from 2 patients, that is 34.5x more records logged from 4x more data sources than state-of-the-art sets. This project aims to get a more accurate disease picture and to reduce the physical and psychological burden of traditional assessment methods. Ultimately, the work has the potential to save patients' time and improve the efficiency and effectiveness of health services.

Problem

PD is a neurodegenerative disorder affecting around ten million people worldwide [13]. Its motor and non-motor symptoms worsen patients' quality of life, especially in the elderly population. Traditionally, the severity of PD is quantified using clinical scales during regular visits to health centres [4]. However, this approach is unsuitable for long-term, periodic monitoring because it is expertise-dependant [17, 22] and prone to recall [6, 14, 19] and cognitive bias [1]. Likewise, brief assessment sessions provide an inaccurate picture of PD as its symptoms fluctuate throughout the day [15, 16]. Thus, a sporadic assessment makes it difficult to tailor treatments to the patient's real condition [20, 23].

Some of these issues are tackled using electronic devices which are objective, more concise and more precise than traditional methods. Nevertheless, the device type, the way it is used and the chosen monitoring methodology have different outcomes. Although wearable devices are a popular choice, they are often attached to uncomfortable body locations and patients need to follow scripted assessment routines. This makes it impractical to monitor PD for a long time outside the laboratory. Recently, inertial sensors, touchscreens and cameras in smartphones have been utilised to assess PD by measuring people's movement patterns [10], hand tremor [2, 8, 9, 12], freeze of gait [15], gait difficulty [12] and upper-limb bradykinesia [18]. Nevertheless, almost all identified smartphone-based projects monitor participants during a single day, performing short scripted tasks under controlled conditions. Although [10] was the exception and participants were followed doing their regular activities, the authors only used one data source (GPS) and only found a suggested relationship to PD clinical scores. We believe this is an idea that can be further explored and improved.

Methodology

Our work monitors PD in a ***longitudinal***, ***naturalistic***, ***non-intrusive*** and ***non-disruptive*** way. Additionally, we follow a ***macro-scale*** approach assessing trends of human activities and habits instead of measuring fine motor movements (micro-scale) which have been the focus of previous research [9, 12, 23]. Furthermore, we propose to combine multiple data sources (***multi-source***) to simplify the quantification of complex behavioural features. Altogether, these six monitoring attributes are unique in the context of PD progression assessment.

Our project explores PD progression assessment using Latent Behavioural Variables (LBV) derived from heterogeneous data. Such data is collected using a smartphone and processed to infer behavioural metrics. We define an LBV as a set of metrics that quantify a human activity or habit. These LBVs can be analysed over an

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

extended period to identify trends and obscure outliers. Thus, the trends' changes can be mapped to the disease's evolution. Due to time constraints and the complexity of this problem, we will focus on finding one LBV and look for more if there is enough time.

To accomplish all this, we propose a four stage methodology for PD monitoring (Figure 1). During the first stage, ***data collection***, each participant receives a smartphone to log their social, environmental and interaction data using all the sensors and interfaces within it. This is complemented by ambient, spatial and other web data sources. We do not require participants to perform any assessment tasks (*non-intrusive*) or to attach the phone to a particular body location or in any orientation whatsoever (*non-disruptive*). We ask participants to carry the device while they follow their daily routine (*naturalistic*) for several months (*longitudinal*). Then in the ***data processing*** phase, raw data is modified, filtered and ranked to reduce the complexity and dimensionality of the original dataset.

The ***data analysis*** stage has two tasks, LBV Identification and "Profile of Living" (PL) generation. In the first one, we carry out a combinatorial analysis of a group of data sources to infer a human activity or habit. Then, we compute a set of metrics that quantify distinctive features of these inferences. For example, if we were analysing the "human mobility" LBV, we could infer the time and location when an individual is running, walking or standing. These inferences would be based on location, spatial, accelerometer, Wi-Fi and Bluetooth data in order to calculate metrics like the duration of the movement sessions, the average walking speed, the maximum travelled distance or the places visited by the monitored person. During the second task, the evolution of the LBV is quantified using a PL. A PL is a proposed concept where each of the LBV's metrics is divided into two. The portion obtained at the beginning of the monitoring period produce a personal baseline while the rest is considered as deviations over time. If we go back to our example, we could analyse each of our metrics (i.e., walking speed) at different time granularities (i.e., daily, weekly, monthly) and decide at what point in time we should split them into two groups based on the chosen time cycle(s). Then, the second group (deviations) can be compared against the first one (baseline) to quantify any differences in the magnitude of a metric and generate one score for each time cycle. This process would be done for each metric of an LBV.

Finally, in the ***evaluation*** step, LBV variations (shifts in behaviour) are mapped to changes in PD severity. This mapping is evaluated using clinical scores produced at regular intervals by trained staff as a ground truth. Such scores will come from the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [4] that assess motor and non-motor symptoms and is widely used as a golden reference in research projects outside clinical environments [11]. Thus, we will study the correlation between the variation scores of the LVB's metrics and the disease's MDS-UPDRS scores and sub-scores. Besides this and as a secondary measure, we will analyse the trends of the different LBV(s)' metrics, to see if their changes are related to each other and to the theoretical progression of the PD symptoms they quantify according to the literature.

We expect two main contributions from this approach. The first one is a methodology to investigate behavioural LBVs related to PD using time series data. The second one is the identification of at least one LBV that, if correlated with PD severity, will be a proof of concept of non-intrusive and non-disruptive PD monitoring based on passive mobile sensing.

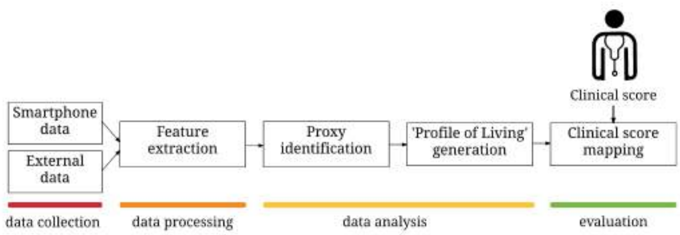


Figure 1. Methodology for PD monitoring using latent behavioural variables.

Results to date

To first test this methodology, we carried out a pilot monitoring study where 29 types of data sources (smartphone's interfaces and sensors and web sources) were logged from two patients over 83 days using the Android app AWARE [3] modified in-house to allow data encryption and cloud synchronisation. The resulting dataset (D1) has ~290 million data points. In comparison to existing smartphone collected datasets for PD monitoring (D2 [5] & D3 [21]), ours has a higher resolution (107,680 records per person per monitored hour (R/P/Hr)) and semantic richness (29 sources). This means, D1 is 34.5x bigger than D3 (2,060 R/P/Hr from 7 sources), and although it is 5x smaller than D2 (359,243 R/P/Hr), the latter has only one source of data. This approach to data collection will increase the potential for inferring complex PD-related behavioural habits.

Next, we identified six LBVs, each composed of several metrics based on the collected data, the symptoms of PD [7], the assessment tasks of the MDS-UPDRS [4] and everyday human activities or habits that might be influenced by PD symptoms according to what other works have measured using alternative approaches (i.e. [2, 10]). These LBVs are typing patterns, phone usage patterns, episodes of going up/down stairs, participant's indoors routine, motor activities, and social patterns. The "social patterns" LBV combines Bluetooth (BT), Wi-Fi, location, spatial, calls, and messages data and has metrics like BT surrounding devices (a potential indicator of human presence) or communication patterns (a possible indicator of a positive mood), among others. We preliminarily analysed the "social patterns" LBV using three days of data from one participant. From the Wi-Fi and BT data, it was possible to identify periods during which the participant was outside the home. Furthermore, spatial and Wi-Fi data was used to infer that the patient visited different urban locations including his/her home. In the BT log, a sporadic device was recorded while the person was at home indicating social interaction with another individual. All these parameters can be weighted and put together to generate a social interaction score. Following a similar process, other LBVs and their metrics can be identified and analysed. The next step is to take into account data from a longer period and study patients' routines to then proceed to evaluate these inferences against the PD clinical scores.

Conclusions

The pilot results provide good evidence that PD progression monitoring based on behavioural inferences extracted from data collected using mobile devices is worth exploring further. Due to the exploratory nature of this work, there are several LBVs that can be analysed, all promising leads that might be related to PD severity. If we find at least one PD-related proxy at the end of this proof-of-concept monitoring methodology, future work under the same line of research could have a significant impact on the quality of life of PD patients by saving them time, reducing the physical and psychological burden related to traditional and alternative assessment methods, and improving the precision of treatments and interventions. This will help to lessen the clinicians' workload and improve the efficiency of health services.

Ethical statement

The pilot study was approved by the NRES Committee North West - Greater Manchester West, REC reference: 15/NW/0043.

References

1. Chaudhuri K.R., Martinez-Martin P., Schapira A.H. V, Stocchi F., Sethi K., et al. (2006). International multicenter pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for Parkinson's disease: the NMSQuest study. *Movement disorders : official journal of the Movement Disorder Society* 21(7):916–23.
2. Daneault J.-F., Carignan B., Codère C.É., Sadikot A.F., Duval C. (2012). Using a smart phone as a standalone platform for detection and monitoring of pathological tremors. *Frontiers in human neuroscience* 6:357.
3. Ferreira J.J., Godinho C., Santos A.T., Domingos J., Abreu D., et al. (2015). Quantitative home-based

assessment of Parkinson’s symptoms: The SENSE-PARK feasibility and usability study. *BMC neurology* 15:89.

4. Goetz C.G., Tilley B.C., Shaftman S.R., Stebbins G.T., Fahn S., et al. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement disorders : official journal of the Movement Disorder Society* 23(15):2129–70.
5. Hammerla N.Y., Fisher J.M., Andras P., Rochester L., Walker R., et al. (2015). PD Disease State Assessment in Naturalistic Environments using Deep Learning. *Conference on Innovative Applications of Artificial Intelligence* (Austin, Texas).
6. Hicks J., Ramanathan N., Kim D., Monibi M., Selsky J., et al. (2010). AndWellness. *Wireless Health 2010 on - WH ’10* (ACM Press, New York, New York, USA), p 34.
7. Jankovic J. (2008). Parkinson’s disease: clinical features and diagnosis. *Journal of neurology, neurosurgery, and psychiatry* 79(4):368–76.
8. Kostikis N., Hristu-Varsakelis D., Arnaoutoglou M., Kotsavasiloglou C., Baloyiannis S. (2011). Towards remote evaluation of movement disorders via smartphones. *Conference proceedings : . Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference* 2011:5240–3.
9. Lemoyne R., Mastroianni T., Cozza M., Coroian C., Grundfest W. (2010). Implementation of an iPhone for characterizing Parkinson’s disease tremor through a wireless accelerometer application. *Conference proceedings : . Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference* 2010:4954–8.
10. Liddle J., Ireland D., McBride S.J., Brauer S.G., Hall L.M., et al. (2013). Measuring the lifespace of people with Parkinson’s disease using smartphones: proof of principle. *JMIR mHealth and uHealth* 2(1):e13.
11. Movement Disorder Society Task Force on Rating Scales for Parkinson’s Disease (2003). The Unified Parkinson’s Disease Rating Scale (UPDRS): Status and Recommendations. 18(7):738–750.
12. Pan D., Dhall R., Lieberman A., Petitti D.B. (2015). A Mobile Cloud-Based Parkinson’s Disease Assessment System for Home-Based Monitoring. *JMIR mHealth and uHealth* 3(1):e29.
13. Parkinson’s Disease Foundation (2015). Coping with a Diagnosis. < http://www.pdf.org/en/newly_diagnosed_pd >. Accessed January 25, 2016.
14. Patel S., Lorincz K., Hughes R., Huggins N., Growdon J., et al. (2009). Monitoring motor fluctuations in patients with Parkinson’s disease using wearable sensors. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society* 13(6):864–73.
15. Pepa L., Ciabattoni L., Verdini F., Capecci M., Ceravolo M.G. (2014). Smartphone based Fuzzy Logic freezing of gait detection in Parkinson’s Disease. *2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA)* (IEEE), pp 1–6.
16. Pérez-López C., Samà A., Rodríguez-Martín D., Català A., Cabestany J., et al. (2015). Monitoring Motor Fluctuations in Parkinson’s Disease Using a Waist-Worn Inertial Sensor. Lecture Notes in Computer Science., eds Cabestany J, Rojas I, Joya G (Springer Berlin Heidelberg, Berlin, Heidelberg), pp 461–474.
17. Piro N.E., Baumann L., Tengler M., Piro L., Blechschmidt-Trapp R. (2014). Telemonitoring of patients with Parkinson’s disease using inertia sensors. *Applied clinical informatics* 5(2):503–11.
18. Printy B.P., Renken L.M., Herrmann J.P., Lee I., Johnson B., et al. (2014). Smartphone application for classification of motor impairment severity in Parkinson’s disease. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp 2686–2689.
19. Shulman L.M., Pretzer-Aboff I., Anderson K.E., Stevenson R., Vaughan C.G., et al. (2006). Subjective report versus objective measurement of activities of daily living in Parkinson’s disease. *Movement disorders : official journal of the Movement Disorder Society* 21(6):794–9.
20. Sung M., Marci C., Pentland A. (2005). Wearable feedback systems for rehabilitation. *Journal of neuroengineering and rehabilitation* 2:17.
21. The Michel J. Fox Foundation (2013). Predicting Parkinson’s Disease Progression with Smartphone Data. < https://www.kaggle.com/c/predicting-parkinson-s-disease-progression-with-smartphone-data >. Accessed November 26, 2015.
22. Tsanas A., Little M.A., McSharry P.E., Ramig L.O. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson’s disease symptom severity. *Journal of the Royal Society, Interface / the Royal Society* 8(59):842–55.
23. Tzallas A.T., Tsipouras M.G., Rigas G., Tsalikakis D.G., Karvounis E.C., et al. (2014). PERFORM: A System for Monitoring, Assessment and Management of Patients with Parkinson’s Disease. *Sensors (Basel, Switzerland)* 14(11):21329–57.

The Effect of Conspecifics on Auks Behaviour on the Colony Surface

Daria S. Vodolazova and Anna V. Klenova

Department of Vertebrate Zoology, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russian Federation. aethia@yandex.ru

Introduction

To study animal communication one should know more about influence of individuals on one another. Recent theory argues that social complexity is the key driving force behind both communicative complexity and social processes like cooperation, aggression etc. Here we tested how the presence of conspecifics could modify behaviour of individuals in different bird species. We studied three colonial seabirds: crested auklet *Aethia cristatella*, parakeet auklet *Cyclorrhynchus psittacula* and horned puffin *Fratercula corniculata*; Auks (Alcidae). These species form large colonies on the cliffs and slopes of the North Pacific islands and show complex social behaviour there. They often form "clubs" on the colony surface, in which groups of birds of one or several species usually spend a lot of time during all breeding period. However studied species differed in the level of social activity, nest density, colony sizes and some other life-history characteristics. Here we analyse the effect of conspecifics on the occurrence of social and non-social behaviours in three species during colony attendance.

Methods

We collected data in June-August 2010, 2011 and 2013 on Talan Island, Sea of Okhotsk where the large colonies of all three studied species are present. We observed for auks behaviour on two study plots from hides during the peaks of birds activity in the morning. We fixed the behaviour of focal birds and their nearest environment in the radius of 0.5-0.7 m for 4-25 min using digital cameras. We analysed 40 selected videos per species where we used as focal animals both individually banded birds (21 crested auklets, 14 parakeet auklets and 7 horned puffins) and also unbanded ones (19, 26 and 33 individuals of each species respectively). Among 40 videos for each species, 20 files were recorded in presence of small amount of conspecifics (0-4), another 20 files - in presence of large amount of conspecifics (5 and more) within 5 corpuses from the focal bird for more than a half of video duration.

Behaviour analysis was conducted in The Observer XT (Noldus Information Technology). We composed ethograms for each species; totally we identified nine types of behaviour that were common for all three species and one type of behaviour that was specific for crested auklet only. We identified three types of self-maintenance behaviour: horizontal posture, vertical posture, self preening. We also identified six/seven types of social behaviour: moving, agonistic behaviour, visual display, duet, duet rejection, acoustic display and tactile display (for crested auklet only).

We estimated the effect of conspecific in close proximity (within 2 corpuses or closer – effect of "neighbour"), and effect of conspecifics at a farther distance (in 5 corpuses) on the occurrences of each behaviour type in each species. For statistical analysis we used Mann-Whitney U-test and considered that the differences were significant if $p<0.05$. We applied also the Bonferroni correction to determine statistical significance of our multiple correlation analyses.

Results

Crested auklets, parakeet auklets and horned puffins spent relatively the same time in presence of "neighbour" (75, 74 and 66% of time on the colony surface respectively, Mann-Whitney U-test, $p>0.05$). With "neighbour", birds of all three species spent more time on social behaviour in total, than they did without it. The occurrences of most of the social behaviour types were also slightly and significantly higher. Some types of social behaviour occurred only in presence of "neighbour" (agonistic behaviour, duet and duet rejection). Occurrences of self-maintenance behaviour types differed slightly in presence and absence of “neighbour”, but there were no general trends for the species (see Table 1).

Table 1. *Mean±SD* values for occurrences of different behaviours in presence and absence of “neighbour” in three auk species. The results of Mann-Whitney U -test for comparisons between values in presence and absence of “neighbour” (U rates / p values) were also given. Values that were significantly effected by presence of “neighbour” (p<0.05) were printed in bold. Asterisk marks values that were still significantly different after applying the Bonferroni correction. N=40 videos for each species.

Types of behaviour		Crested auklet		Parakeet auklet		Horned puffin	
		With neighbour	Without neighbour	With neighbour	Without neighbour	With neighbour	Without neighbour
Self-maintenance behaviour	Horizontal posture	0.03±0.13	0.01±0.25	0.02±0.08	0.00±0.01	0.11±0.20	0.14±0.30
		706.0 / 0.366		818.0 / 0.567		768.0 / 0.758	
	Vertical posture	0.42±0.32	0.46±0.38	0.65±0.25	0.46±0.41	0.69±0.31	0.61±0.41
		792.5 / 0.942		691.5 / 0.088		788.0 / 0.908	
	Self preening	**0.11±0.17**	**0.05±0.13**	0.01±0.16	0.01±0.04	0.02±0.06	0.02±0.06
		481.0 / 0.002		841.0 / 0.714		780.0 / 0.847	
Self-maintenance behaviour in total		0.57±0.31	0.52±0.39	0.67±0.23	0.47±0.41	0.82±0.31	0.76±0.39
		758.0 / 0.686		701.5 / 0.106		671.5 / 0.216	
Social behaviour	Moving	0.05±0.04	0.14±0.22	0.05±0.07	0.08±0.18	0.02±0.03	0.03±0.06
		791.5 / 0.935		751.5 / 0.243		758.0 / 0.686	
	Agonistic behaviour	**0.01±0.01**	**0.00±0.00**	**0.02±0.02**	**0.00±0.00**	0.00±0.00	0.00±0.00
		499.0 / 0.004		**462.0 / 0.000***		740.0 / 0.564	
	Visual display	**0.25±0.24**	**0.00±0.00**	**0.14±0.2**	**0.00±0.00**	**0.04±0.03**	**0.00±0.00**
		205.0 / 0.000*		**291.5 / 0.000***		**417.0 / 0.000***	
	Duet	**0.10±0.12**	**0.00±0.00**	**0.08±0.13**	**0.00±0.00**	0.02±0.06	0.00±0.00
		380.0 / 0.000*		**485.5 / 0.000***		620.0 / 0.083	
	Duet rejection	0.00±0.01	0.00±0.00	0.01±0.04	0.00±0.00	0.00±0.00	0.00±0.00
		640.0 / 0.124		775.5 / 0.341		700.0 / 0.336	
	Acoustic display	0.01±0.03	0.02±0.06	0.02±0.08	0.04±0.09	0.01±0.07	0.00±0.02
		779.5 / 0.844		868.5 / 0.904		779.0 / 0.840	
	Touch displays	**0.01±0.02**	**0.00±0.00**	-	-	-	-
		370.0 / 0.000*		-		-	
Social behaviour in total		**0.43±0.31**	**0.16±0.22**	**0.33±0.23**	**0.12±0.19**	**0.07±0.13**	**0.04±0.08**
		353.0 / 0.000*		**393.5 / 0.000***		**533.0 / 0.010**	

With large amount of conspecifics in 5 corpuses crested and parakeet auklets spent more time on social behaviour in total, than they did with small amount of conspecifics (43.5% and 35.4% instead of 38.3% and 25.9% respectively), whereas horned puffin showed opposite trend (5.9% and 12.3% respectively). However all these differences were non-significant (Mann-Whitney U-test p>0.05 for all comparisons). Also the amount of conspecifics in 5 corpuses effected the occurrences of only few behaviour types significantly in all species. But all these differences became non-significant after applying the Bonferroni correction. Depending of species, occurrences of some behaviour types were higher, whereas occurrences of other ones – were lower with large amount of conspecifics (see Figure 1). We also found that focal individuals stayed with "neighbour" more often when they were surrounded by large amount of conspecifics.

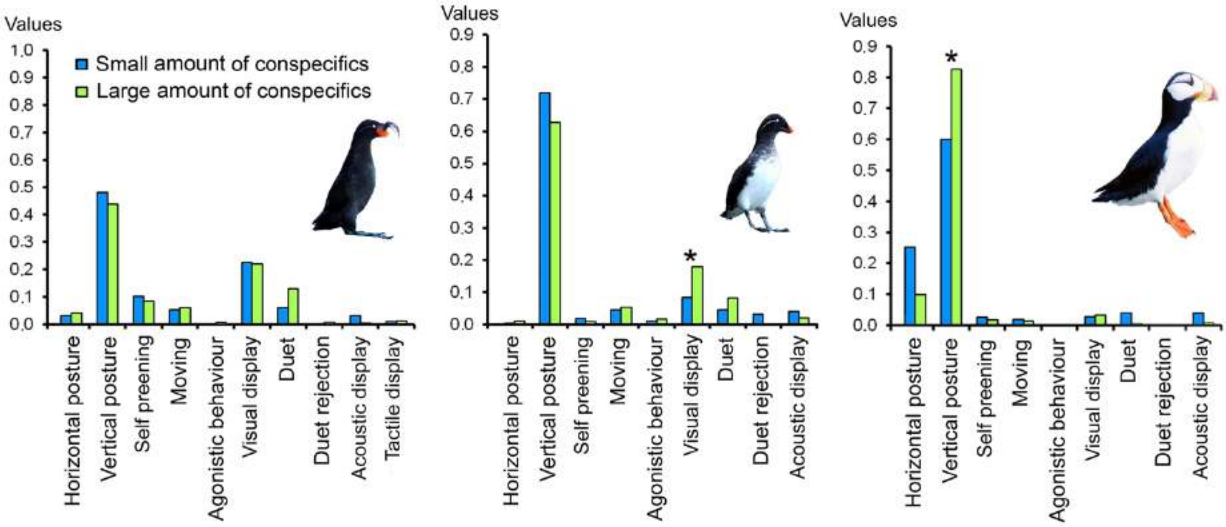


Figure 1. The occurrences of behaviour types in crested auklet, parakeet auklet and horned puffin with small and large amounts of conspecifics in 5 corpuses. N=40 videos for each species. Asterisk marks significant differences (Mann-Whitney U -test, $p<0.05$; however these differences became non-significant after applying the Bonferroni correction).

Conclusion

We found that the presence of conspecifics in close proximity to the focal bird had a considerable and similar impact on behaviour of all three species. With such "neighbour" all birds spent more time on social behaviour in total; occurrences of some types of social behaviour were also remarkably increased. Some interspecific variability in this effect can be partly explained by biological characteristics of the species, such as sizes and density of their colonies and "clubs", which are lager and denser in crested auklet, but smaller and sparser in parakeet auklet and horned puffin. The conspecifics at farther distance had little effect on behaviour of the birds.

Ethical statement

No specific permissions were required for our study location or species according to sections §44 and §6 of the Federal Law of the Russian Federation No. 52 from 24.04.1995 (last update 07.05.2013) "On Wildlife". There were no Special Protected Natural Territories in our study area, and study species are not listed in the Red List of the Russian Federation. Our activities did not include withdrawal of investigated species from nature. Data collection protocol #2010-42 was approved by the Committee of Bio-ethics of Lomonosov Moscow State University.

Acknowledgements

This study was funded by the Russian Scientific Foundation (grant 14-14-00237).

Altered Open Field and Nocturnal Homecage Locomotor Activity in the Olfactory Bulbectomized (OB) Rat Model of Depression

Z.McAleavey[1,2] and J.P.Kelly[1,2]

1Discipline of Pharmacology, School of Medicine, NUI Galway, Galway, Ireland.z.mcaleavey1@nuigalway.ie
2Galway Neuroscience Centre, NUI Galway, Galway, Ireland

Animal models are an invaluable tool in understanding both the causality and the treatment of many disorders. The OB rat model is a well validated model of depression and involves bilateral removal of the olfactory bulbs. The resulting behavioural and neurochemical alterations are reflective of the human condition, with hyperactivity in the open field (OF) being the primary behavioural hallmark of this model. The OF (see Figure 1a) is a round circular arena, 75cm in diameter, with reflective aluminium walls (60cm high) and is illuminated to a degree that the animals find aversive (over 200 lux). Thus the hyperactivity exhibited by OB animals in the arena is thought to be due to the animal's inability to cope with a novel or stressful environment – a behaviour that is attenuated by chronic but not acute antidepressant treatment [1] which is reflective of the clinical scenario. However, although stress is an element of some depressive episodes it is not common to all. Nocturnal homecage activity (HCA) is another parameter that can be assessed (see Figure 1b), where the animal's locomotor activity in a familiar environment can be monitored [2]. We have developed a tracking system that enables repeated evaluation of this parameter and avoids any exposure to a novel/stressful environment. Using these two behavioural parameters, further insight can be gained as to whether the hyperactivity in the OF is a maladaptive response to stress or if hyperactivity is an inherent characteristic of this model.

The tracking system developed in our lab consists of 3 x 24-place home cage racks (72 places in total) with an individual camera above each cage, allowing the entire cage floor to be seen. A white plastic insert is placed underneath the food hoppers in each cage to prevent the animals going underneath and thus the animal is visible at all times. In order to be able to visualise the animals' nocturnal activity, free standing infrared lights (RS Components, IR illuminator with 24 LEDs) are positioned on the floor opposite the cage racks (0-5 lux at cage level). The free standing infrared lights do not rely on beam breaks of infrared photobeams, instead they offer diffuse background lighting, thus illuminating the entire cage. The video continuously recording (24-hrs) onto a central digital video recorder (DVR) allows the animal's nocturnal HCA to be tracked at a later date using EthoVision tracking software. To our knowledge, many of the current HCA monitoring systems available rely on breaking infrared photobeams, without accompanying video. An advantage to the system developed in our lab is that not only does it offer the opportunity to assess nocturnal HCA but the video recording allows specific behaviours (e.g. stereotyped) to be concomitantly evaluated if desired.

Male Sprague-Dawley rats were bred in house, pups used in the study were weaned at PND21 and remained group housed until they were singly housed one week prior to surgery (approx. 7 weeks old). Animals were housed in standard plastic bottomed cages containing 3Rs Lab© bedding, in a temperature controlled room (20-24°C), relative humidity (45-65%) and 12:12h light dark cycle (lights on at 08.00h). Rats were fed a standard laboratory diet of rat chow pellets, with food and water available *ad libitum.* The rats (approx. 8 weeks old) underwent sham and OB surgery and were left to recover for 14-days. Surgery groups were randomly assigned by body weight, with no significant difference in body weight between groups prior to surgery (p>0.05). 24-hr HCA was recorded for the sham and OB animals from the day before and every day for the duration of the study. In the morning of the 14-th day post-surgery the animals were tested in the OF, after which nutritional enrichment was provided once weekly for the remainder of the study. OF locomotor activity and nocturnal HCA at baseline (night before surgery), day 0 (day of surgery), day 7 and day 13 post surgery were scored at a later date using EthoVisionXT 8.5© tracking software. All locomotor activity was measured as 'Distance Moved (cm)' therefore accounting for the total distance moved by the rat in the arena. Exposure to the OF was for 5min and locomotor activity was measured for the full duration, starting when the animal was placed in the centre of the arena. The nocturnal HCA activity was measured for the duration of the rat's nocturnal cycle. All video files – including OF and HCA were recorded onto a central DVR system. One OF area is recorded per channel, however 8 home cage places are recorded onto one channel on the DVR system thus allowing 8 home cage places to be tracked simultaneously.

OF locomotor activity data satisfied the criteria for parametric analyis and thus was analysed using an independent t-test, with p<0.05 being deemed significant. Nocturnal HCA data did not satisfy the criteria for parametric analysis and therefore non-parametric equivalents were used. Using SPSS, a Friedman's ANOVA followed by Mann-Whitney *post-hoc* (where appropriate), with p<0.01 being deemed significant. Animals with incomplete bulb removal or frontal cortical damage were removed from analysis, resulting in 35 sham and 29 OB animals.

The experimental protocol was carried out in accordance with the guidelines and approval of the Animals Care and Research Committee, National University Ireland, Galway, under licence from The Health Products Regulatory Authority (HPRA).

OB animals were significantly more hyperactive compared to their sham counterparts in the OF on the 14th day post-surgery (t(62) = 3.35, p<0.01). In the HCA, there was a significant difference over time between the groups (Fr=57.04, p<0.001), there was no significant difference between sham and OB animals at baseline (U=386 p>0.05) or on the night of surgery (U=432, p>0.05). However, 7 days (U=232 p<0.001) and 13 days (U=122, p<0.001) post surgery OB animals were significantly more hyperactive compared to the sham animals.

The above results suggest that the hyperactivity in the OB animals may not be environmentally specific and therefore may be an inherent characteristic of the model, rather than a transient maladaptive characteristic. This is further supported by our findings that the hyperactivity is evident 7-days post surgery and persists at least to the night before (day 13) OF testing. Further analysis is needed to assess how soon after surgery the hyperactivity becomes evident and if this affected by chronic antidepressant treatment, in a similar fashion to that observed in the OF. If this proves to be the case, then the HCA paradigm will be a useful measure of time of onset of antidepressant effects in this model.

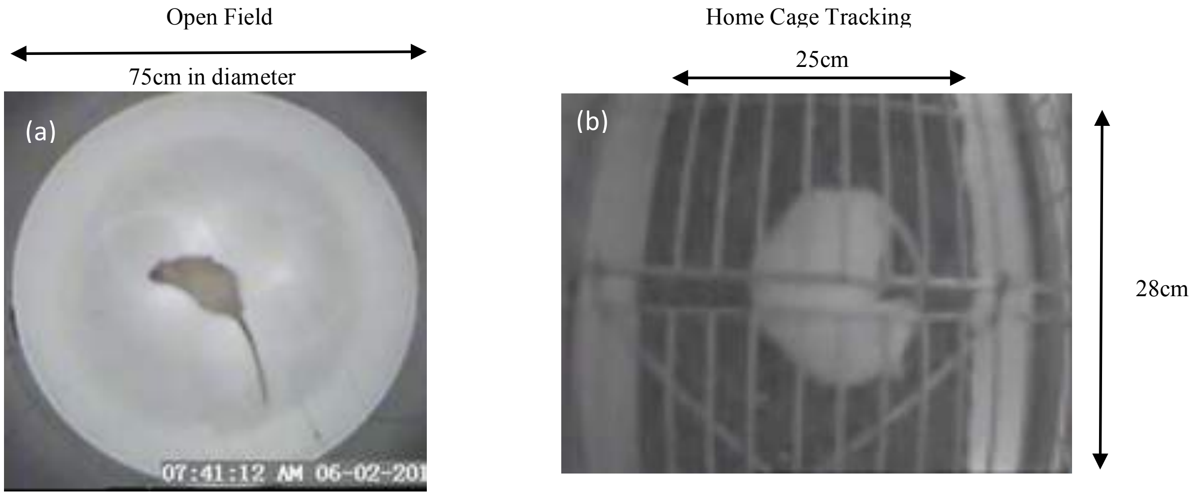


Figure 1. Open field and home cage apparatus. (a) Open field (OF) (b) Home Cage Tracking

Acknowledgments

Special thanks is given to Ambrose O'Halloran for the design and construction of the 72-place HCA apparatus.

ZMcA is a recipient of a NUI Galway Hardiman post-graduate scholarship

References

1. Roche, M., Harkin, A., Kelly, JP. (2007) Chronic fluoxetine treatment attenuates stressor-induced changes in temperature, heart rate, and neuronal activation in the olfactory bulbectomized rat. *Neuropsychopharmacology*, 32:1312-1320
2. Dunne, F., O'Halloran, A., Kelly, JP. (2007) Development of a home cage locomotor tracking system capable of detecting the stimulant and sedative properties of drugs in rats. *Progress in neuro-psychopharmacology & biological psychiatry.* 31:1456-1463

Implantable Hearing Devices in Sheep

Martijn Agterberg[1,2], Frank Walboomers [3], Andy Beynon [1], Theo Peters [1], Joost Zwartenkot[1], Rishubh Verma[4,] Daniel Smyth[4,] and Emmanuel Mylanus[1]

[1]Department of Otorhinolaryngology, Donders Institute For Brain, Cognition and Behaviour, Radboudumc, Nijmegen, The Netherlands
[2]Department of Biophysics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands
[3]Department of Biomaterials, Radboud University Nijmegen, The Netherlands.
[4]Cochlear, Mechelen, Belgium

Objective

The developments of acoustic implant technology continues and evolves, and the need for a large animal model is increasing. In the present study we investigate whether the sheep is suitable for testing relatively large human hearing implants. Both the osseointegration and the functionality of the hearing implants are investigated.

Material and method: Skulls were used for implantation of bone-conduction hearing implants, cochlear implants, and the Codacs®. After implantation animals were housed for twelve weeks at the farm of the animal facilities. Functionality of the implant was tested in awake animals, however the most straightforward test of functionality, namely detection of a behavioural response after stimulation with the implant, was challenging. Auditory brainstem responses were recorded in anesthetised and awake animals. Correct device placement was verified by cone beam CT and histology. Histological sections of the osseointegrated bone-conduction device were produced with a modified sawing microtome technique. All animal experimental procedures were approved by the University's Committee on Animal Research (RU-DEC 2014-172).

Results

Two year old sheep are a good in vivo model for both bone-conduction devices and Codacs®. The sheep middle ear and cochlea are, although smaller, highly similar to the human middle ear and cochlea. No signs of infection of the large titanium implants and subcutaneous tube were recorded. All screws of the implanted hearing devices were osseointegrated within the temporal bone. With the sawing microtome technique relatively large titanium implants (>2 cm) could be processed. Auditory brainstem responses could easily be recorded in awake animals.

Conclusion

Older sheep are valuable as animal model for hearing implants, especially when larger (human) implants are tested. ABRs can be measured in awake sheep. For in vivo implantation of Codacs we advise to sacrifice the eardrum and part of the bony-ear-canal, otherwise it is hard to visualize the positioning of the implant on the oval window.

References

1. Schnabl et al., 2012; Ter Haar et al., 2010; Eeg-Olofsson et al., 2014

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Use of the Visual Cliff apparatus for high-throughput quantification of impulsivity

Price Dickson, Summitt Liu, Troy Wilcox , Tyler Roy, Kathryn McNaughton, Laura Anderson and Elissa Chesler

The Jackson Laboratory, United States of America

Addiction is a critical public health issue for which the underlying genetic mechanisms remain largely unknown. Behavioral traits including impulsivity, risk-taking, the response to novelty, and anxiety are associated with addiction related behaviors in humans and experimental animals. We and others have shown that these behavioral associations are driven by shared genetic mechanisms and have begun to leverage the genetic overlap among addiction and associated behavioral traits to identify genes driving addiction using a high throughput approach. For years, this approach has been amenable to identification of genes underlying novelty and anxiety traits because indexing these traits is relatively simple, rapid, and cost effective in mice and rats. Specifically, anxiety-related behavior, exploratory behavior, hyperactivity, and response to novelty can be measured in a single session in an open-field type apparatus. In contrast, characterization of impulsivity is a relatively time consuming and expensive process. Currently, the most commonly used impulsivity screen is the 5-choice serial reaction time task which requires a 3 – 4 month testing protocol using an operant conditioning chamber. In order to facilitate the prospective identification of genes underlying impulsivity, it is necessary to create a high throughput, non-invasive, and inexpensive impulsivity assay. To that end, we have evaluated the use of the Visual Cliff apparatus to index impulsivity in the laboratory mouse. The Visual Cliff is similar to an open field with two exceptions. First, the apparatus is suspended three feet above the ground by cantilevering it over the edge of a laboratory bench covered with a checkered plastic table cloth that is draped onto the floor below. The apparatus floor is transparent and is constructed of clear Plexiglas, enabling mice to detect the height difference between the countertop and floor. This arrangement gives the appearance of a cliff. The apparatus is monitored overhead by a video camera and the test is scored using the Noldus Ethovision system. When mice are placed in the apparatus they immediately move to the counter-top side. Many mice then make rapid and repeated sorties onto the "cliff" side of the apparatus, which change in frequency and duration over the testing session. To characterize the behavioral constructs measured in this apparatus, we have devised multiple quantitative endpoints and used them to evaluate the behavioral and neurobiological correlates of visual cliff behaviors in a genetically diverse population. We tested male and female mice from 64 BXD strains on the Visual Cliff apparatus. We calculated heritability of Visual Cliff traits in the BXD panel to assess the reliability of the phenotypic endpoints among genetically similar individuals. Following this, we used GeneNetwork to identify genetic correlations among Visual Cliff traits and other behaviors, notably retrieving behavioral and neurochemical indicators of alcohol and drug use. Collectively, these findings show covariation among behaviors in the Visual Cliff and behaviors and genetic mechanisms associated with drug use. These data provide evidence that Visual Cliff exploration, which varies with genotype, represents a form of impulsivity which can be measured in a high throughput manner.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Automatised Movement Analysis in the Inverted Grid Test: Validation using the MPTP Mouse Model of Parkinsonism

E. Palasz[1], W. Niewiadomski[2], M. Skupinska[1,2], M. Zylinski[3], M. Steczkowska[1], A. Gasiorowska1[,2], G. Niewiadomska[1], G. Riedel[4]

[1]Laboratory of Preclinical Studies in Neurodegenerative Diseases, Nencki Institute of Experimental Biology, Warsaw, Poland. {g.niewiadomska@nencki.gov.pl}
[2]Department of Applied Physiology, Mossakowski Medical Research Centre, Warsaw, Poland. {wniewiadomski@imdik.pan.pl}
[3]Institute of Metrology and Biomedical Engineering Department of Mechatronics, Warsaw University of Technology, {Warsaw, Poland. zapomnialem@o2.pl}
[4]School of Medical Sciences, University of Aberdeen, Aberdeen, Scotland. {g.riedel@abdn.ac.uk}

Background

One of the big challenges in behavioural neuroscience is the robust detection of motor deficits when concerned with experimental models of movement diseases. Automated recordings can help in standardizing the recording and may aid in the analysis of multiple parameters. Utilisation of gait analysis tools (for example: CatWalk - Noldus IT, Wageningen, The Netherlands; DigiGate - MouseSpecifics, Inc. Framingham, MA, USA; TreadScan – CleverSys, Reston, VA, USA) has led to more robust and repeatable studies. However, such tools are widely missing for sensory-motor analysis tools such as the 'traction or inverted grid test'. Typically, a mouse or rat is placed onto a grid and this is subsequently turned upside down such that the subject is now clinging onto it. The principle endpoint is the taken as overall cling time (time to fall off), but further movement-related parameters could be analysed. Automatic detection of paws movement in this task has proved difficult, so a human observer has to be employed with the consequence that results heavily depend on subjective rating. Consequently, between and within laboratory rating may proof highly variable and a more standardized recording and analysis tool is warranted. We here set out to develop a more objective method, whereby the role of experimenter has been limited merely to the identification of paws and their position. From clearly identified paws position, we then created a Matlab (The Matworks) script to enable the auto-detection of movements of individual paws with a high fidelity. Furthermore, data enabled the calculation of both static and dynamic movement parameters and were validated using a pharmacological model of Parkinson's disease.

Method

Animals and drug groups: Male, 3-month-old, C57BL/6 mice (Medical Research Centre Animal Farm, Bialystok, Poland), weighing 25-30 g at the beginning of the study were used and remained sedentary throughout the test. Mice were housed eight to ten animals per cage with food and water available *ad libitum*, at constant temperature and humidity (23±1^{o}C, 55±5%) on 12-h light-dark cycle (lights on: 8:00 am). A total of 10 animals were used in this study. Animal husbandry was provided by the Nencki Institute animal care facility. All experiments were conducted with the approval of the local ethics commission (Polish Law on the Protection of Animals) and carried out in accordance with National Institute of Health's Guide for Care and Use of Laboratory Animals (Publication No. 85-23, revised 1985) and the European Communities Council Directive (63/2010/EU). The acute experimental 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) schedule was applied (Kurosaki et al. 2004). The MPTP is a neurotoxin widely used to produce experimental models of PD. Six C57BL/6 mice randomly selected from two cages were injected four times with MPTP hydrochloride (20 mg/kg in saline, intraperitoneally; Axon Medchem, Cat.No.1075) at 2-h intervals within a day. Four saline treated C57BL/6 mice served as control. The inverted grid test was performed on days 10 and 20 post-surgery and tissue harvest commenced after the second behavioural measurement.

Inverted grid test - Apparatus and recording: A dark grey wire mesh grid (12 x 12 cm) with 5 mm^2 openings surrounded by opaque Perspex walls, 9 cm high, was utilised. Each mouse was placed in the center of the grid and it was turned upside down so the mouse was hanging and clinging to the grid. The distance between grid and the floor was 24 cm and soft padding was provided to mitigate the falling down. The maximal 'cling time' was set to 30 seconds during which subjects were freely moving on the underside of the grid. Each test session (on day 10 and on day 20 post-treatment) consisted of 3 trials with an inter-trial interval of 1 minute. An overhead CCTV camera (Sony α37) recorded the trials at 25 frames per second and Mpeg files were written to the hard drive of a PC running Windows 7.

Movement detection and auto-analysis using Matlab script: Since each test session consisted of 3 trials, our analysis concentrated on the trial with the longest total Cling Time. Videos of the selected trial of each mouse were imported into our Matlab script (MouseTrac) with a user interface for manual tagging of video items frame-by-frame. After calibration to the external reference frame of the grid (12x12 cm) for the extraction of positional data (all given in mm), the rater blind to the animal's treatment tagged each individual paw (front left, FL; Front Right, FR; Hind Left, HL; Hind Right, HR), paw position and the start and the end of grip. Our MouseTrac script enabled continuous tracking of all paw markings simultaneously over time (30 seconds; frame by frame) with subsequent storage of raw data and export to Excel. Paw movement was further fractionated into three principal components: Hold, Step with or without touch(es), and Wall contacts (for definitions, see Table 1). Primary metrics that were extracted for all behavioural components included timing (start, stop) as well as the number of occurrences. The automatic analysis of MouseTrac provides highest accuracy with respect to onset and termination of specific movement-related events and average timings. From these primary readouts, secondary proxies were accessible (also defined in Table 1).

Statistical analysis: All data were expressed as mean ± SD and compared by ANOVA using within and between subject factors followed by multiple range Newman-Keuls post-hoc test. The analysis was made using STATISTICA 10 software at $P < 0.05$ level of significance.

Table 1: Definition of primary proxies from the MouseTrac system. Secondary proxies are derived from these data and can be calculated using for example Excel. Note that most of these parameters can be individually calculated for each paw.

MouseTrac parameters	Excel parameters	Definition
Cling Time		**Overall time in which animal kept contact with a grid (latency to fall; max 30s)**
Hold		**Time with all digits seizing the grid wire and XY coordinates being constant**
Hold Time [s]		Total time in Hold during the trial
Number of Holds		Total number of Hold episodes during the trial
	Mean Hold duration [s]	Hold Time/Number of Holds
	Hold Index [n/s]	Number of Holds/Cling Time
	Hold Time Index	Hold Time/Cling Time
Step		**Paw displacement from given XY coordinates of a given Hold to new XY coordinates of the next Hold; a Step may include a Touch (see below)**
Step number		Number of completed steps during the given trial
Step duration [s]		Cumulative time spent in Step mode during trial
Step length [mm]		Cumulative distance of all Steps during a given trial
	Mean Step duration [s]	Step duration/Step number
	Mean Step length [mm]	Step length/Step number
	Mean Step velocity [mm/s]	Step length/Step duration
	Step Index [mm/s]	Step length/Cling Time
Touch		**During Step, a paw makes contact with grid; however the phalanges do not perform a grip but the paw moves on to a different XY coordinate or returns to its original location (exploratory steps with touches)**
Number of touches		Total number of touches during trial
Touch duration		Total time performing touches
	Touch Index [n/s]	Number of Touches/Cling Time
	Step with touches Index	Number of Steps with Touches/Cling Time
Step with touches		Number of Steps with one/multiple touches
	Step with touch Index	Number of Step with touches/ number of Steps
Wall contact		**During Step, a paw makes contact with the wall**
Number of wall contacts		Total number of wall contacts per trial
Wall contact duration		Total time in wall contact
	Wall contact Index	Number of Wall Contacts/Cling Time
	Wall contact time Index	Sum of durations of Wall Contacts/Cling Time
Step with wall contact		Number of Steps with wall contact
	Wall contact Index	Number of Wall Contacts/number of Steps
	Wall contact time Index [s]	Duration of Wall Contacts/Step duration
Train		**Series of displacements (>2) of the same forepaw**
Number of Trains		Number of occurrences of trains
	Steps per Train	Number of Steps in Train/Number of Trains
	Train Index	Number of Trains/Cling Time
	Regularity Index	Regularity of alternation between long and short Steps during movement

Results

From the unbiased detection of forepaw activity we extracted a movement – time histogram (Fig. 1A) to describe the complex sequence and extend of clinging movements.

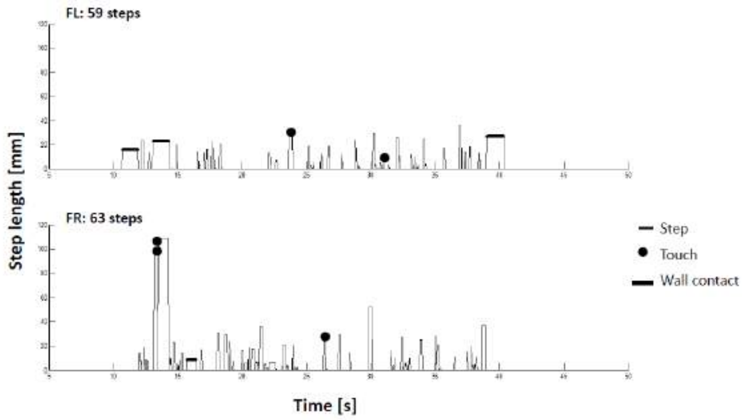


Fig.1A. The example of two step-grams, for left and for right forepaw in the same mouse in the same trial are shown. Following features of movement on the inverted grid may be discerned: alternating use of left and right forepaw, short series of consecutive steps performed with the same paw (named in this paper trains of steps), and alternation of steps' length. In this example also episodes of Wall Contacts and Touches took place – there were two Steps with single Touch performed by left forepaw, one Step with double Touch and one Step with single Touch performed by right forepaw.

Results for each paw was quantified and revealed some novel aspects of movement. For example, each paw was engaged in only very few successive steps. There was a positive correlation between step length and velocity with which these were executed and a robust alternation between long and short steps of any given paw emerged (Fig. 1B).

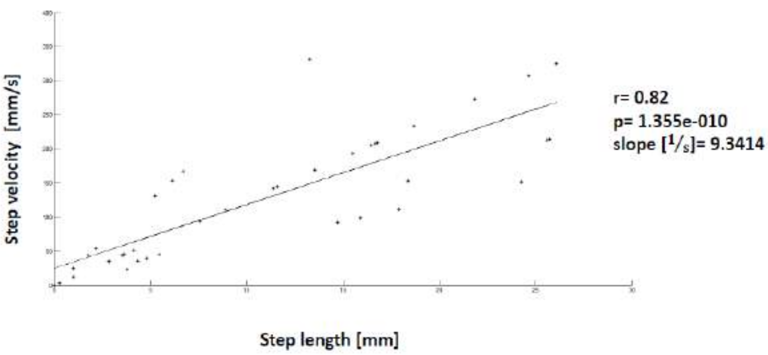


Fig.1B. Changes in step length are accompanied by changes in step velocity; long steps are performed with higher velocity than short ones. The correlation between step length and step velocity: the 9.34 1/s slope calculated for the example presented means that in average 10 mm increase in step length would be connected with increase in velocity of 93.4 mm/s.

There were also some differences in motor performance between control and parkinsonian mice. MPTP mice performed significantly more forepaws steps and distance covered by forepaws was longer in comparison to controls. Of particular note was that control mice hardly ever perform paw movements that include touches; by contrast, a significant increase in steps with touches especially for the front paws was noted in the MPTP cohort (Fig. 2).

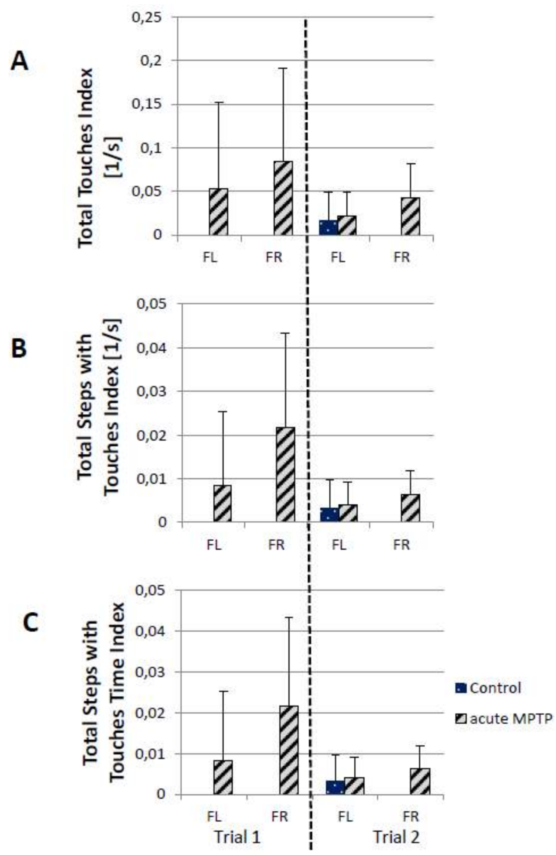


Fig.2. Indices derived from detailed analysis of forepaws movement in C57BL/6 saline and MPTP treated mice in the Trial 1 (10 days after treatment) and in Trial 2 (20 days after treatment) of inverted grid test. **(A)** Total Touches Index representing the number of touches performed in the unit of time by a forepaw was statistically different between controls and MPTP. **(B)** Total Steps with Touches Index expressing number of steps, within which at least one touch was performed was far greater in MPTP mice than in controls. **(C)** Total Steps with Touches Time Index showing share of cling time which fells to steps with touches was statistically higher in MPTP in comparison with control group.

This detail would have completely escaped the standard analysis method relying on overall cling time (not different) or counting the number of steps (not different either). The overall richness in data analysis parameters of this novel tool is capable of revealing even subtle but more realistic and detailed movement patterns differences between test cohorts and facilitate translation to clinical conditions.

Conclusions

Collectively, we here describe a first attempt to standardize the analysis of the traction test/inverted grid test and introduces parameters of high sensitivity to neural interruption. Clearly that heterogeneity of data reported on for example experimental models of Parkinson's disease have not provided confidence into existing pharmacological of genetic model organisms. However, one reason could be the relative crudeness of our behavioral standards and we propose that by developing more sensitive analysis tools will also increase a) reproducibility of results and b) lead to greater translation of proxies.

This work was funded by NCN grant 2014/15/B/NZ4/05041

Justice Stories: Measuring Content and Emotion in Employee Interviews About Fairness in the Workplace

Karl N. Kelley[1], Emilia A. Knapczynski[1], Katherine M. Fermin[2], Amanda L Conlin[3]

[1]North Central College, Naperville, IL USA
[2]FlavorChem Corporation, Downers Grove, IL USA
[3]Northern Illinois University, DeKalb, IL USA

Introduction

This project was designed to field-test two competing theoretical models of organizational justice. The three primary goals for this project included (1) creating an interview that would encourage the participant to tell their story in sufficient detail for analysis; (2) developing a coding scheme that would capture important content elements related to current research in organizational justice, and (3) evaluating the emotions experienced or reported by participants during the interview.

Methods

To create the interview, we utilized a critical incidents procedure suggested by Klimoski & Wilkinson [3] to illicit actual stories of perceived workplace injustices. In this process, the initial questions were broad and open-ended. If the participant struggled in organizing their story, several specific prompts were used to help them focus and provide sufficient detail for analysis. Some examples of the prompted questions were:

- Please describe, in as much detail as you can, the events that led up to the situation, the actual situation itself, and why you believe you were not treated fairly.
- What was it about this situation that made it feel unfair to you?
- What could the manager have done to make this a better situation?

Specifically, we were interested in their understanding of the injustice situation and how they responded to this situation at work and outside of work. All of these interviews were recorded.

To analyze the content of the stories, we focused on the theoretically suggested elements of organizational justice. This included distributive justice (resource outcome), procedural justice (following clear rules and policies), informational justice (being kept informed of the process and reasons for the decision), and interpersonal justice (social treatment focusing on dignity and respect). Two coders listened to each story and reported the different justice elements they heard. We organized these responses using a crystallization process suggested by Ellingson [4]. This process begins with creating a broad conceptual impression of an overall story then identifying, organizing, and interpreting patterns of specific elements (i.e. the types of organizational justice observed in this project).

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

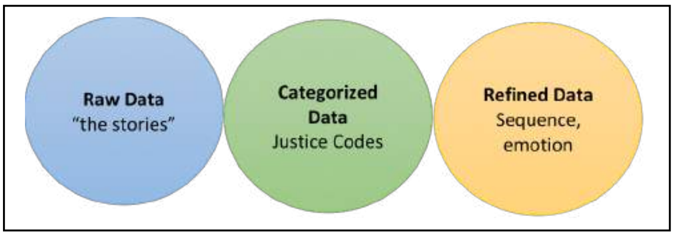


Figure 1: The coding process

To analyze the emotional content of these stories we used The Affect Grid developed by Russell, Weiss, & Mendelsohn [5]. This is a single response mood scale that requires participants to indicate a current mood by placing a mark on a 6 X 6 grid. The grid has two dimensions: pleasure–displeasure (horizontal) and arousal–sleepiness (vertical). In the current study, coders listened to the interviews and marked their perception of the participants' stated or experienced emotional state at regular intervals.

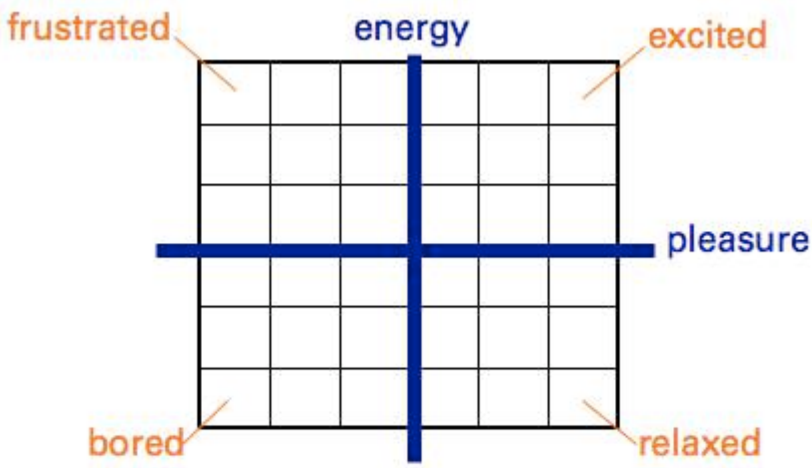


Figure 2: The Affect Grid

Results

We are still in the process of analyzing our data. Currently, we have 9 complete injustice stories that range in time from 10 to 45 minutes. An initial analysis was performed via a frequency analysis. Each instance of each type of justice issue was noted several times within the interviews. The results show that the most commonly reported injustice is procedural injustice, at 40 instances followed closely by interpersonal injustice at 35 instances. Seeing the dominant themes of procedural and interpersonal issues, this may suggest that perhaps the more important types of justice within the context of the workplace are procedural and interpersonal justice. In addition to the frequency of the statements, we will be assessing when the statements are made in the story. We are just beginning the affect coding portion of this research.

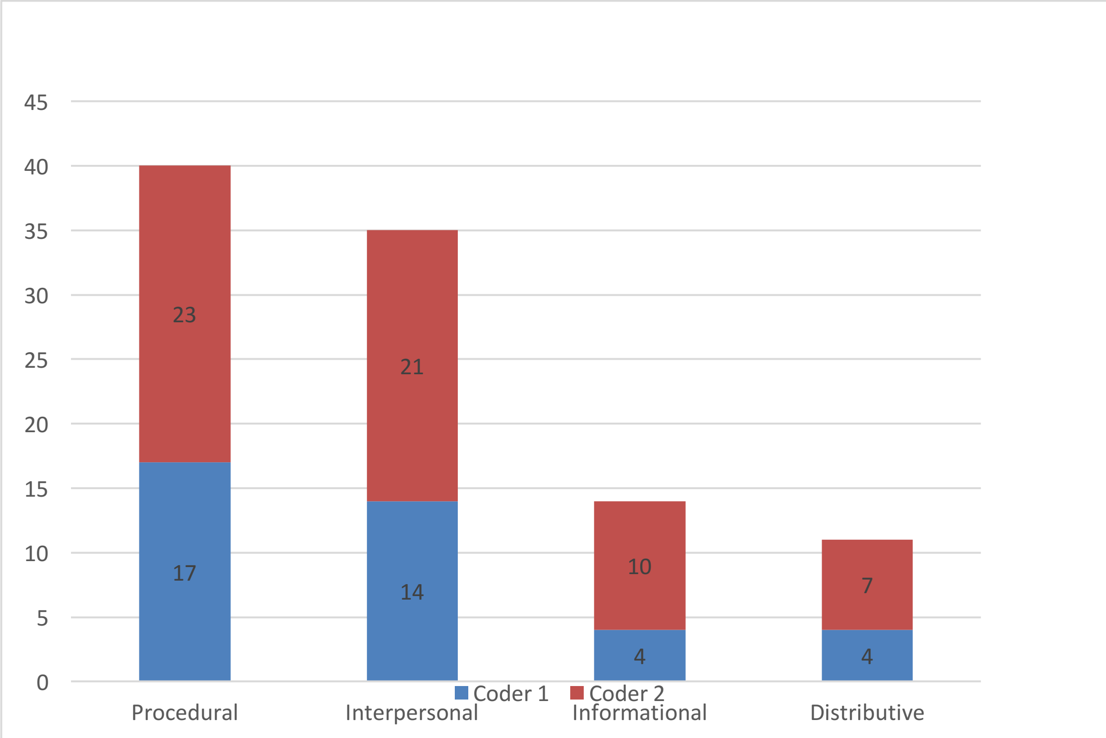


Figure 3: Justice Code Frequency

Issues related to the Measurement of Behavior

The current research raises multiple issues relevant to the measurement of behavior. For example, in the data collection phase, we are concerned about the extent to which the critical incident method yields accurate and valid stories. Specifically, we wonder if the prompts foster accurate recall or create memories? In the analysis phase, there are multiple questions about coding both the content and emotion in the stories. Although we are testing for interrater reliability, we wonder if the coder training is influencing their perceptions by imposing a structure rather than hearing what the participants actually wanted to say. We will discuss the strengths and limitations of these approaches as we seek to develop new ways to measure how perceptions of injustice influence behavior.

Ethics Review

All procedures were approved by the North Central College Research Ethics Committee, September 2014.

References

1. Ambrose, M. L., & Schminke, M. (2009). The role of overall justice judgments in organizational justice research: A test of mediation. Journal of Applied Psychology, 94(2), 491-500.
2. Greenberg, J. (2009). Everybody Talks About Organizational Justice, But Nobody Does Anything About It. *Industrial and Organizational Psychology*, 181-195.
3. Klimoski, R. J., & Wilkinson, T. R. (2013). Psychological assessment in industrial/organizational settings. In J. R. Graham, J. A. Naglieri, I. B. Weiner, J. R. Graham, J. A. Naglieri, I. B. Weiner (Eds.), *Handbook of psychology, Vol. 10: Assessment psychology (2nd ed.)* (pp. 347-372). Hoboken, NJ, US: John Wiley & Sons Inc.

4. Ellingson, L. (2009). Introduction to Crystallization. In *Engaging crystallization in qualitative research: An introduction* (pp. 1-17). Thousand Oaks: Sage Publications.
5. Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493-502.

Parking Safety Architecture - Measuring Perceptions of Safety

M.D. Bredero

Architect and researcher at Saxion University of Applied Sciences
Roggestraat 2, 7411 EP Deventer, The Netherlands
mail@architectmaartendouwebredero.com

Keywords: urban planning, architecture, parking space, comfort zones, privacy, orientation, perception, customer safety

Introduction

Car parking remains a recurrent issue in cities. Several solutions to this challenge have been developed since the introduction of the mass vehicle.

In order to minimize the visual and spatial impact of parked cars, they are either camouflaged or literally hidden. Camouflage in the shape of greenery or sub-terrain positioning is widely used in parking sites around buildings. However, in both cases, cars parked adjacent to the building remain in sight [1]. Within a city fabric, parking garages provide another solution. Either as a free standing building or underground car park. Parking cars on the top of a building is not often utilized. The latter approach obviously has many advantages, such as no risk of inhaling toxic smoke during a fire, double use of the loaded roof, and they are also relatively low in cost. But perhaps disadvantages such as reduced façade surface area and the reluctance of drivers to take on a slope play a major role [2].

In this research proposal, a representative shopping mall of less than 10,000 square meters in rectangular plan (golden section ratio 1 : 1.61) is taken as a reference. The ramp lies adjacent to the longitudinal side of the complex in order to reduce its steepness.
This is of interest in terms of the psychological perception of safety. In this set-up, while taking the common adjacent possibility as a reference, the 'roof' and 'underground' variants will be compared. And one can even imagine a shopping mall with both types of parking facilities. The reduced rental space will in any case outweigh the favourable effect on the perception of the public space around the building. Cars parked in squares and parking lots hamper the appreciation of architectural surroundings and prevent other and more attractive activities, such as street markets or open air concerts, from taking place.

Next to the reference design, actual parking solutions in the City of Deventer will be part of the study.

Research questions

The questions to be considered are:

- How do customers parking their car perceive safety?

- How does general visibility (overview and being seen) improve the perception of safety?

- To what extent does Wayfinding contribute to the perception of safety?

- Does having an overview of the activities going on in and around a building invariably improve perceptions of safety?

- To what degree do entrances/ramps in parking solutions such as underground car parks and roof access hinder perceptions of safety?

- To what degree does architectural space with natural light enhance perceptions of safety?

- To what degree does protection from strong winds enhance perceptions of safety?

- To what extent can a virtual design model correlate with the actual physical building.

Type of research

The research involved is empirical with an inductive approach. Based on specific virtual reality observations of behaviours, an attempt will be made, on the one hand, to deduce general rules. On the other hand, it is assumed, as a deductive hypothesis, that the presence of natural light will enhances feelings of safety. This also applies to protection from strong winds, although this has more to do with being comfortable. Also, having an overview of the spatial surroundings generally seems to be appreciated.

The research will consist of questionnaires being administered at specific intervals, while respondents (novice and experienced drivers) are exploring a to be created virtual environment. A survey will therefore be combined with a case study in order to test the hypothesis.

For instance, the respondents have to answer questions on how safe and smooth (on a scale of one to five) they perceive the space they are virtually wandering through. Or how they appreciate exiting the complex after parking their car and entering it to retrieve their car after shopping [3]. Recording their behavior during the simulated process of passing through and exploring the virtual architectural space, will provide an insight into the psychological perception of spatial information. In this process, the respondents' physical condition, i.e. their eye movements, perspiration and heart rates are measured to register any corresponding level of anxiety. Depending on the number of measurements, the research will primarily be qualitative in nature. Only limited statistical calculations will therefore be necessary.

In the real life situation only questionnaires will be at stake.

Research methodology

Precise measurements can be achieved by working with two distinct groups, comprising:

1) regular visitors to parking areas near shopping malls;

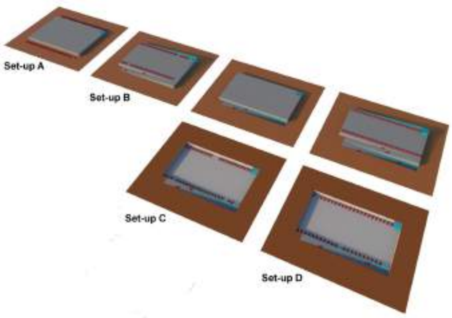
2) novice users of the three parking solutions.

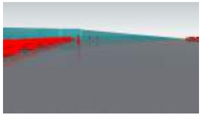
Within a given time frame, the subjects within each group are to explore the further to be refined four virtual models of the shopping mall and respond to specific questions put to them. Set-ups A, B, C and D will then be examined and compared. The comparison will depend on the data gathered.

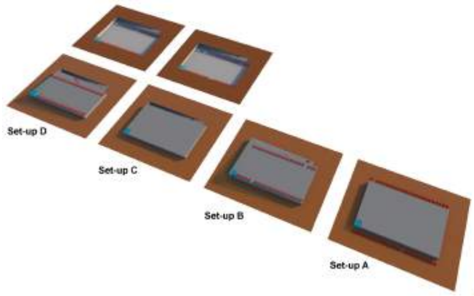
Set-up A, adjacent parking; Set-up B, roof parking; Set-up C, underground parking; Set-up D, a combination of B and C.

Moreover, relevant questions will be put regarding comfort and orientation – i.e. the mental condition of subjects before and after the tests – with the answers being compared with their actual behaviour [4].

The same applies to the groups visiting realized parking garages adjacent to shopping facilities.







Figures above: Four shopping mall designs A-D, covering footprints (including ramps) of 9600 square metres, with a ground level height of 5 meters. The transparent main entrance is positioned diagonally.

Centre figures: Visualization of natural day light underneath (Set-ups C and D) and guiding wind shields on top of the building (B and D).

Note: two opposite main entrances as circulation columns, instead of the present single entrance, will improve internal customer movement and facilitate more escape routes.

Architect Maarten Douwe Bredero & SAXION Centre for Urban and Environmental Developn

Implementation of research

The current architectural Preliminary Design (see below) will be converted further into a semi-realistic environment using advanced visualization and gaming techniques.

Subjects will explore the parking environment – either individually or in groups – by means of computer interfaces. They will interact with computer screens while sitting on chairs at gaming consoles. This can be done, for instance, at the advanced serious gaming facilities at the University of Twente (T-Xchange).

This non-direct method (registration of mental decisions) will allow the mental process of safety assessment to be measured.

General constraints

Realistic model (architectural design preliminary model to be converted into real-life visualization).

Representative measurements (large enough number and representative groups of subjects).

Feasible (usable protocol for professional architects as a result).

Representative parking garages in real life situation.

References

1. Kahlert, D. and Schlicht, W. (2015) Older People's Perceptions of Pedestrian Friendliness and Traffic Safety: An Experiment Using Computer-Simulated Walking Environments. *Int. J. Environ. Res. Public Health.*
2. Medat, C. and Wogalter, M. (2003) Perceptions on Parking Facilities: Factors to consider in Designs and Maintenance. Proceedings of Human Factors and Ergonomics Society 47th Annual Meeting.
3. *Stark, J. (2012) Parking Lots: Where Motorists Become Pedestrians* <http://www.albany.edu/ihi/files> Accessed Janauary 15th , 2016.

What is Trending in Eye Tracking Scanpaths on Web Pages?

Sukru Eraslan[1, 2], Yeliz Yesilada[2], Simon Harper[1] and Alan Davies[1]

[1] School of Computer Science, University of Manchester, Manchester, United Kingdom
sukru.eraslan@postgrad.manchester.ac.uk | simon.harper@manchester.ac.uk | alan.davies-2@manchester.ac.uk
[2] Middle East Technical University Northern Cyprus Campus, 99738 Kalkanli, Guzelyurt, Mersin 10, Turkey
seraslan@metu.edu.tr | yyeliz@metu.edu.tr

Eye Tracking on the Web

In this paper, we introduce a new concept called trending scanpath that represents the most popular scanpath among users.

Eye tracking has commonly been used to analyse user behaviours on the Web [2]. When users navigate on web pages, they make a series of fixations. In other words, their eyes become relatively stationary at certain points. These series define their scanpaths on the pages. Figure 1 illustrates an example of a user scanpath on the Godaddy page that is segmented into its visual elements such as menu items and search boxes [1, 4]. The circles represent the fixations where the sizes of the circles are directly proportional the durations of the fixations. Besides, the sequence is represented by the numbers inside of the circles. By tracking eye movements of users, we can investigate which parts of web pages are popular among the users and which paths they follow to complete their tasks.

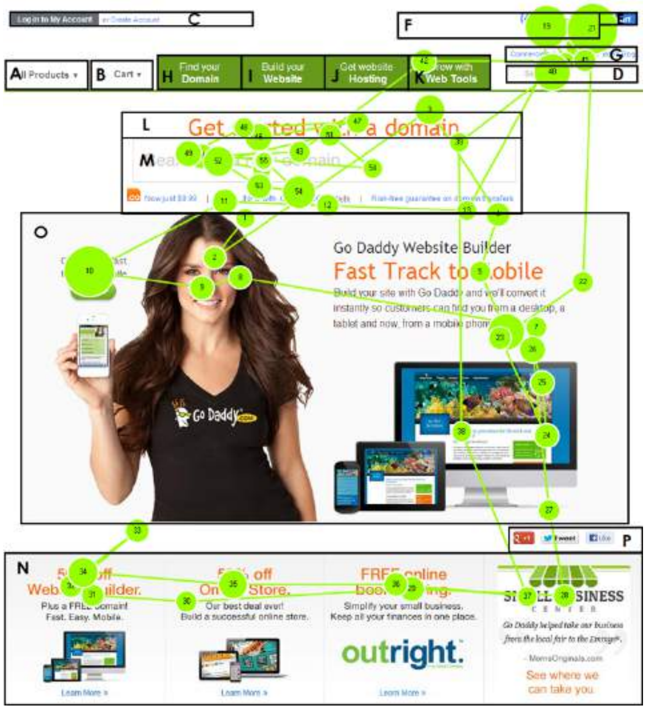


Figure 4. A scanpath of a particular user on the Godaddy page

In order to analyse how users interact with visual elements of web pages, their scanpaths should firstly be represented in terms of the visual elements of the web pages. In particular, if a user firstly fixates the element O, secondly the element F, thirdly the element L and finally the element M on the Godaddy page, his or her scanpath should be represented as OFLM.

A number of different techniques have been proposed and/or applied to analyse scanpaths in the literature [3]. These techniques have mainly been used for (a) determining similarities/dissimilarities between a pair of scanpaths, (b) computing transition probabilities between the elements of a web page, (c) detecting patterns in multiple scanpaths, and (d) identifying a common scanpath of users on a particular page by combining their scanpaths [3]. In this paper, we introduce a new concept in combining multiple scanpaths into a single scanpath called trending scanpath.

Trending Scanpath

Trending scanpath is the most popular scanpath on a particular page among users. The main features of this scanpath are as follows:

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

- The trending path shows **a general direction** which is followed by most users to complete their tasks. Existing techniques try to identify a path shared by all users as a common scanpath [3]. However, these techniques are likely to lose some shared elements in intermediate levels of processing due to a variety of the positions of the elements in the scanpaths. Therefore, the shared elements may not be included in their resulting scanpaths. Because of this reason, these techniques have a reductionist approach which means that their resulting scanpaths can be unacceptably short (or nothing) to understand user behaviours. In contrast, the trending scanpath includes both the elements shared by all users and other elements which get at least the same attention as the shared elements in terms of the total duration of fixations and the total number of fixations on the elements.

- The trending scanpath is incrementally **developed** by adding new users. When there is no user, the trending scanpath is empty. As illustrated in Figure 2, after a certain point, the scanpath becomes stable and that is the one we call trending path [4].

- The trending path is developed based on the current content of elements and can be **changed** as the content of the elements change. For example, an element may include an article which is about a popular news and the element may be included in the trending path. However, when the content of the element is changed, it may not attract attention of users anymore, as a consequence it may not be included by the trending scanpath.

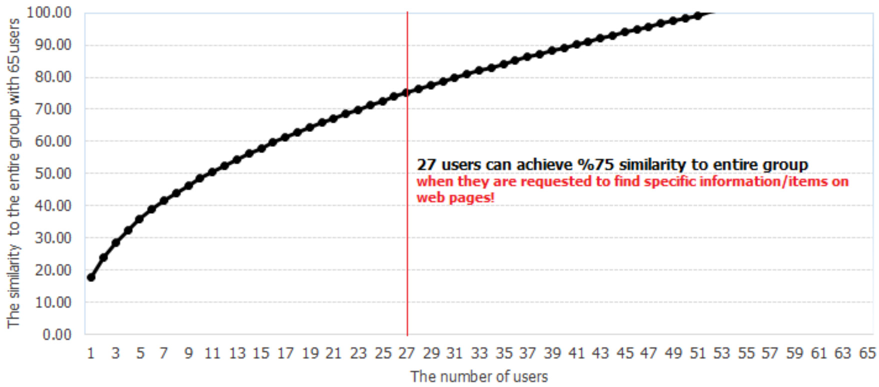


Figure 5. The similarities between the results of the sub groups of 65 users to the entire group [4]

As the Oxford English Dictionary[1] defines *Trend* as "**a general direction** in which something is **developing** or **changing**", we believe that the trending scanpath term is the most appropriate term here. In the literature, when scanpaths are combined into a single scanpath, the scanpath is typically referred to as a common scanpath. According to the Oxford English Dictionary[2], one of the definitions of *Common* is "shared by, coming from, or done by two or more people, groups, or things". Hence, the common scanpath term by itself may lead us to understand that there is only one path and it is shared by all users. Therefore, this term is not suitable for representing our concept here because there is no restriction to include only shared elements.

As illustrated in Figure 3 [4], to identify a trending scanpath, trending elements should be identified by analysing user scanpaths. In other words, we should identify which visual elements should be in the trending scanpath. These elements should then be combined based on their overall positions in the user scanpaths.

Specifically, a trending scanpath should be comprised of trending visual elements. When a particular visual element is shared by all user scanpaths, it should be identified as a trending element. Even though existing techniques focus on the shared elements, they tend to lose them because of their positions [3]. For example, when the user scanpaths O**FLM** and **FL**N**M** are available the Goddady page, FLM cannot be discovered as a

[1] http://www.oxforddictionaries.com/definition/english/trend
[2] http://www.oxforddictionaries.com/definition/english/common

pattern by the *eyePatterns* scanpath analysis tool because of the element N [5]. However, the trending scanpath should include the shared elements regardless of their positions. Other elements with at least the same attention as the shared elements should also be identified as trending elements. The trending elements should then be positioned in the trending scanpath according to their overall positions. For example, when a particular element is firstly fixated by most users, it should be positioned at the beginning of the trending scanpath.

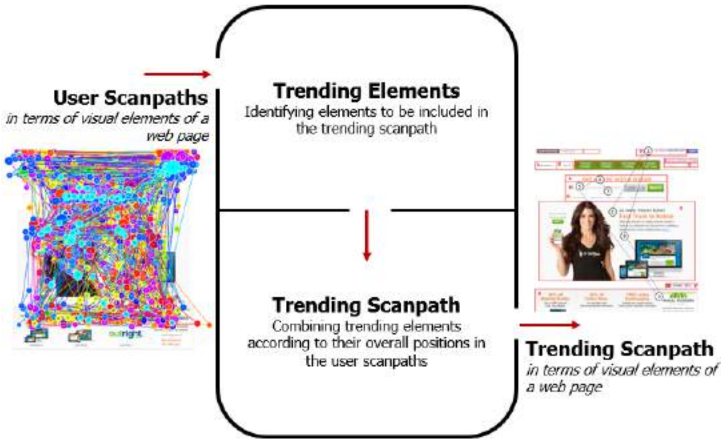


Figure 6. A concept to identify a trending scanpath of users in terms of visual elements of a web page.

Contributions of Trending Scanpaths

We introduce a trending scanpath in this paper. It contributes to user behaviour research on the Web by addressing the weaknesses of existing research [3]. It could be used for different objectives. Specifically, a trending scanpath on a particular web page can be counted as a guide to transcode (namely, re-engineer) the web page to make it more accessible in constrained environments such as on small screen devices and in audio representation [6]. By making the firstly and commonly visited visual elements more accessible, small screen device users can access these elements without the need of a lot zooming/scrolling and visually disabled users can access these elements without spending a lot of time on unnecessary things. This is known as "Experiential Transcoding" in the literature [6].

References

1. Akpınar, M. E. and Yeşilada, Y. (2013). Vision Based Page Segmentation Algorithm: Extended and Perceived Success. *In Current Trends in Web Engineering (LNCS), Sheng, Q. Z. and Kjeldskov, J. (Eds.)*, Vol. 8295. Springer International Publishing, 238-252.
2. Ehmke, C. and Wilson, S. (2007). Identifying Web Usability Problems from Eye-Tracking Data. *In Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It,* Vol. 1, Swinton, UK, British Computer Society, 119-128.
3. Eraslan, S., Yesilada, Y. and Harper, S. (2016). Eye Tracking Scanpath Analysis Techniques on Web Pages: A Survey, Evaluation and Comparison. *Journal of Eye Movement Research,* 9, (1):2, 1–19.
4. Eraslan, S., Yesilada, Y. and Harper, S (2016). Eye Tracking Scanpath Analysis on Web Pages: How Many Users? *In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, ETRA 16*, New York, NY, USA, ACM, 103-110.
5. West, J. M., Haake, A. R., Rozanski, E. P. & Karn, K. S. (2006) eyePatterns: Software for Identifying Patterns and Similarities Across Fixation Sequences. *In Proceedings of the 2006 ACM Symposium on Eye Tracking Research & Applications, ETRA 06,* New York, NY, USA, ACM, 149-154
6. Yesilada, Y., Harper, S. and Eraslan, S. (2013). Experiential transcoding: An EyeTracking approach. *In Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*. ACM, 30.

Stress Interface Inducer, a Way to Generate Stress in Laboratory Conditions

J-M. Diverrez, N. Martin and N. Pallamin

Uses and Acceptability Lab, b<>com, Cesson-Sévigné, France. jean-marc.diverrez@b-com.com

Introduction

Behavioral studies carried out in the laboratory are often confronted by the issue of generating stimuli for cognitive states. This study proposes a framework for stimulating stress in the laboratory in order to collect data under controlled conditions.

Stress research may be divided into two main categories (acute vs. chronic) related to the temporality of the stressor [1]. Acute stressors are the ones which are presented for a duration of some minutes while chronic stressors are presented for weeks [2] Our paper concerns the elicitation of acute stress under laboratory conditions. In order to justify the different choices made for the conception of the stress interface inducer (SII), a short literature review concerning the theories and the methodologies used in stress researches is initially provided. Then, the development of SII and the relevant protocol are described. Finally, the results of one experiment are presented.

1. Stress: literature review

The versatility of the term stress is so wide that some authors like Stoke and Kite [3] argued about its usefulness as a scientific term. Tepas et al. [4] pointed out for instance that the term stress is often associated with a variety of constructs like: adaptation, anxiety, arousal, burnout, coping, exertion, exhaustion, exposure, fatigue, difficulty, mental load, repetitiveness, strain, stressor, and tension. The vagueness of the term stress is a consequence of its adoption in a variety of research domains that range from organizational studies to psychiatry. Not surprisingly, even at a theoretical level, no agreement can be found concerning the phenomenon of stress. In search of simplicity the different theories are often grouped into 3 main categories: the response-based models, the stimulus-based models and the interactional models. Response-based models see stress as a cluster of psychological and physiological responses in reply to a challenging situation. This approach focuses on the consequences of the stress and is represented by authors like Selye [5] who introduced the notion of General Adaptation Syndrome to describe the three stages of the body resistance to prolonged stress. Stimulus-based models, on the other hand, focus on the events that cause a stress response. In particular, they analyze the characteristics that a stimulus must possess in order to provoke stress. Three of these main characteristics are overload, conflict and uncontrollability. Stimulus-based researchers like Ivancevich and Matteson [6] have studied the influence of limited time and high performance standard on work overload. The transactional models see stress as a unbalance between environmental demands and individual resources. According to this view, stress responses are created when a threat is perceived and the individual is unable to cope with it. Probably the most influential model of the transactional approach is the one proposed by Lazarus [7]. This model is characterized by two stages of appraisal. In the primary appraisal, the subject evaluates if the situation represents an actual threat (i.e. is "relevant to" and is "conflictive with" the individual's goals). In the secondary appraisal, the individual evaluates the resources available to face the threat and decide the coping strategies.

2. Stress Measurements

In order to identify the theoretical framework and the way to elicit stress in our participants, we focused on the subjective and objective measurements of stress. Concerning the subjective measurements of stress, we decided to adopt two standard questionnaires used in stress studies (see Method - *Subjective measurement)*. Regarding the objective measurements, a significant numbers of articles investigating the links between stress and physiological responses have demonstrated that provocation of punctual stress causes physiological responses in individuals. Such responses may be modifications and variability of heart rate, modifications breathing rate,

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

blood pressure and galvanic skin activity [8]. For example, Shi *et al.* [9] shows a strong correlation between stress levels and electrodermal activity (EDA). Pickering *et al.* [8] shows that blood pressure increases at the same time as stress. Healey *et al.* [10] shows a correlation between breathing rate and stress levels. Sierra De Santos *et al.* [11] shows the relevance of measuring stress by measurements of electrodermal and heart activity (95% recognition rate). Lastly, Partala *et al.* [12] demonstrates that there is a link between pupil activity (dilation and constriction) and stress.

Method

1. Material

The variety of theoretical frameworks that investigate stress is reflected in the amount of paradigms adopted to elicit the feeling of stress in a laboratory. A multitude of stressful tasks has been proposed ranging from simple tracking tasks [13] to more complex methods like the Montreal Imaging Stress Task [14] or software like GASICA [2]. Some of these tasks tend to produce stress exploiting social reaction. A typical example of such task is the “Trier Social Stress Test” [15] in which participants are asked to perform a short speech in front of an audience. Other experimental paradigms rely on the unbalance between task and resources to cause stress. The most common paradigms which are used to obtain such unbalances are based on the presence of a secondary interfering task [16] or on temporal constraints. This last was the option chosen to provoke a stress response in our participants. Our experimental design, in fact, is based on the procedure proposed by Campbell [17] to investigate the effect of time on simple mathematical operations. Thereby, we created stressful situations where individuals had to carry out additions under time pressure. 3 conditions were created:

- *Condition 1 - Time to response is large*: participants had to answer following a beep sound occurring 2650ms after the calculation was presented.
- *Condition 2 - Time to response decreases across trial*: participants had to answer before a beep sound. This beep occurred 2650ms after the calculation was presented for the first trial and was reduced to 900ms for the last trial. At each trial, time before the beep sound decreased by 50ms.
- *Condition 3 - Time to response is short*: participants had to answer before a beep sound and the beep occurred 900ms after the calculation was presented.

For each experimental condition, there were 36 trials (3 conditions: 108 trials per participant). Each trial consisted of a simple mathematical sum, such as “2+7” or “5+8”. Lastly, to expose the stressful situation to participants, we developed a software program with the following characteristics: display user instructions, displays randomly additions with several response times (i.e. level of time pressure), makes a beep at each trial after a delay fixed by preset conditions (i.e. level of time pressure) and saves performance data (success/error and response times) for each trial.

2. Physiological measurement

The following physiological indexes were measured: cardiac, respiratory and electrodermal responses, and eye-tracking data. Biopac Bionomadix MP150 was used to measure physiological responses. SMI glasses 2 60Hz were used to record eye tracking data.

We used the following indexes to test the effect of conditions on physiological data: electrodermal activities (EDA), heart rate activities (ECG RR / ECG R Wave) and respiration activities (Respiratory Rate). For each index, we computed the mean by condition.

3. Subjective measurement

After each condition (i.e. 36 trials), participants filled out two standardized scales to evaluate the effect of induced stress on subjective feeling: The first questionnaire is the Short Stress State Questionnaire (SSSQ) [18] that evaluates 3 aspects of the feeling of stress (Engagement, Distress and Worry). The SSSQ is actually a simplified version of the original Dundee Stress State Questionnaire [19]. The second questionnaire is the Raw-TLX (RTLX), a simplified version of the NASA Task Load Index [20] that has proved to perform comparably to the original version [21]. The RTLX assesses the perceived workload of a task as a simple (unweighted) sum of 6 dimensions (mental demand, physical demand, temporal demand, performance, effort, and frustration). The choice of these 2 questionnaires allows us to assess different aspects of the feeling of stress in our participants.

4. Participants

24 participants took part in the study and received in exchange a coupon for €15. All participants signed up with informed consent before beginning the experimental procedure and were informed about the goals of the study, procedures, cautions and ethical issues for the participation in the study.

5. Procedure

The following procedure was used during the experiment (see figure 7): before starting calculation, a baseline for physiological measurement is recorded. After, participants start calculation and pass all the 3 conditions (within-subject design). Between each condition, a break is observed to reduce stress levels. To avoid order effect, the presentation of condition is counterbalanced and the presentation of the calculations is randomized.

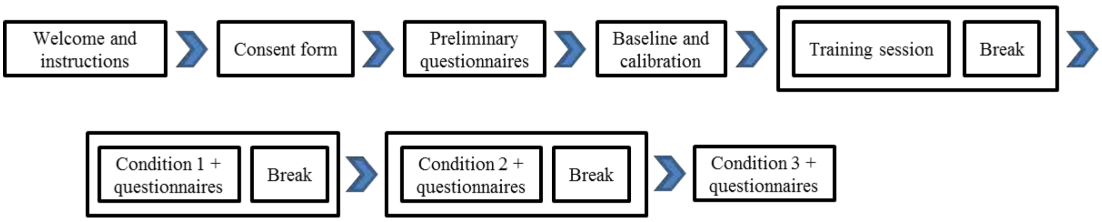


Figure 7. Experimental procedure

Results

We tested the influence of the experimental conditions (i.e. condition 1: long response time, condition 2: response time reducing over time and condition 3: short response time) on subjective and physiological data. Since all conditions were presented to each participant (within-subject design), we used a statistical method that takes into account non-independence to analyze data: a mixed model (also called random effect model) [22]. Mixed models are used because they offer more flexibility and robustness in modeling than repeated ANOVA [23]. Technically, to take into account individual variability and correlation between data, flexibility on intercept is added by including a random parameter (u_{0j}) in the regression equation (see equation 1). Thereby, variability between participants (e.g. personal sensibility to mental exigence) is integrated into the modeling. Since the condition variable (treated with dummy variables) in our experiment is categorical (i.e. the 3 experimental conditions), this equation (see equation 1) is used in our analyses for each dependent variable (DV). In equation 1, Y corresponds to the DV. Subscript j corresponds to the participant and subscript i to an observation nested in a participant. Coefficient β_0 correspond to the intercept (i.e. corresponding to condition 1) and u_{0j} to the coefficient applied to β_0 for each subject. β_1 and β_2 correspond to coefficients activated for conditions 2 and 3 (since the modalities of variable condition are treated as dummy variables). ε_{ij} corresponding to the term error for observations and participants.

Equation 1

$$Y_{ij} = (\beta_0 + u_{0j}) + \beta_1 Condition2 + \beta_2 Condition3 \ + \varepsilon_{ij}$$

Figure 8 gives an illustration of used mixed model. We plotted the prediction of a regression including a random intercept.

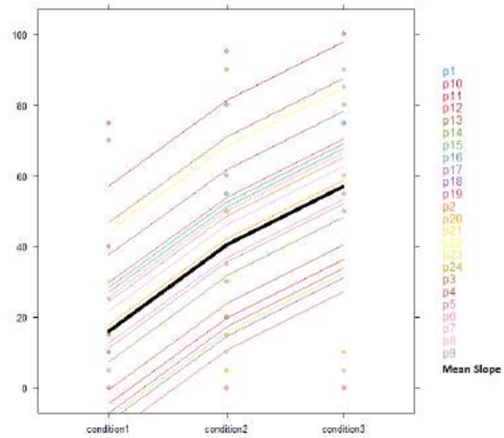


Figure 8. Plot of used mixed model. Bold line indicates the mean effect and colored lines indicate predictions for each participant.

For assessing the effect of a variable in the mixed model, comparison of models is used [24]. Technically, we computed a first model including only random effect (to take into account correlation between data acquired from the same individual). After, we computed a second model including the variable condition. We compared their deviances: these results are presented in Table 2 and Table 4. Lastly, we computed multiple comparisons between conditions.

1. Subjective measurement

Table 1 presents descriptive statistics for subjective measurement. Comparisons of models showed significant differences (see table 2) for all dimensions of RTLX. More precisely, comparison between conditions indicated significant difference between condition 1 (i.e. long response time) and condition 2 (i.e. response time decreasing over time) but also between condition 1 and condition 3 (i.e. short response time) for all RTLX dimensions. For SSSQ, only distress is evaluated as significantly different between conditions.

Table 1. Descriptive statistics (Mean and Standard Deviation) for subjective measurement

Scale	Variable	Condition 1		Condition 2		Condition 3	
		Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
RTLX	Mental demand	16.46	19.86	40.62	29.09	57.29	36.77
	Physical demand	13.54	17.16	30.21	29.06	38.54	32.01
	Temporal demand	22.50	26.91	66.67	19.21	78.12	19.10
	Effort	18.75	23.69	55.00	26.25	68.96	24.14
	Performance	21.25	19.80	50.62	23.83	66.46	23.01
	Frustration	19.38	15.90	39.17	21.35	60.42	22.89
SSSQ	Engagement	30.12	5.79	30.17	4.84	29.38	4.75
	Worry	13.04	4.64	12.17	4.39	13.79	5.35
	Distress	10.00	1.67	13.71	4.33	16.29	6.48

Table 2. Comparison of models between conditions

Scale	Variable	Chi-square	p-value	Condition 1 - Condition 2	Condition 1- Condition 3	Condition 2- Condition 3
RTLX	Mental demand	χ^2 (2) = 35.73	<.001	*z* = 4.24***	*z* = 7.17***	*z* = 2.93**
	Physical demand	χ^2 (2) = 19.81	<.001	*z* = 3.26**	*z* = 4.89**	*z* = 1.63 NS
	Temporal demand	χ^2 (2) = 52.75	<.001	*z* = 8.44***	*z* = 10.63***	*z* =2.19 NS
	Effort	χ^2 (2) = 55.56	<.001	*z* = 7.81***	*z* = 10.82***	*z* = 3.01**
	Performance	χ^2 (2) = 46.59	<.001	*z* = 5.79***	*z* = 8.91***	*z* = 3.12**
	Frustration	χ^2 (2) = 54.55	<.001	*z* = 4.76***	*z* = 9.87***	*z* = 5.11***
SSSQ	Engagement	χ^2 (2) = .92	.34	*z* = .05 NS	*z* = -.95 NS	*z* = -1.00 NS
	Worry	χ^2 (2) = .72	.40	*z* = -1.00 NS	*z* = .86 NS	*z* = 1.85 NS
	Distress	χ^2 (2) = 30.03	<.001	*z* = 3.75***	*z* = 6.35***	*z* = 2.61**

*Signifiant codes: ***: p<.001; **: p<.0.01; *: p<.0.05; NS : Non-Signifiant*

2. *Physiological measurement*

Table 3 presents descriptive statistics for physiological measurement. Comparison of models showed significant differences (see table 4) for EDA and Respiratory Rate. More precisely, significant differences appear between condition 1 and 2 but also between condition 1 and 3 for EDA and Respiratory rate. Currently, the valuable eye-tracking data is not presented in this document, but it will presented in a future paper.

Table 3. Descriptive statistics for subjective measurement

Variable	Condition 1		Condition 2		Condition 3	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
EDA	.47	.03	.49	.04	.49	.04
ECG RR	.35	.26	.32	.28	.34	.27
Respiratory Rate	.34	.24	.26	.22	.24	.18
ECG R Wave	.32	.24	.31	.26	.33	.26

Table 4. Comparison of model between conditions

Variable	Chi-square	p-value	Condition 1- Condition 2	Condition 1- Condition 3	Condition 2- Condition 3
EDA	χ^2 (2) = 7.70	<.001	*z* = .02 *	*z* = .01 *	*z* = -.31 NS
ECG RR	χ^2 (2) = 1.21	.54	*z* = -1.08 NS	*z* = -.41 NS	*z* = .66 NS
Respiratory Rate	χ^2 (2) = 19.60	<.001	*z* = -3.63 ***	*z* = -4.56 ***	*z* = -.93 NS

ECG RWave	χ^2 (2) = .75	.69	z = -.63 NS	z = .18 NS	z = .81 NS

Conclusion

Capacities of SSI to generate stress have been confirmed by subjective and objective measurements. Results show significant differences between low stress conditions and high stress conditions, for all dimensions of RTLX, and for the distress dimension on SSSQ. Differences were also found between the stress conditions on physiological data (ECG RR and Respiratory Rate).

Futures works will focus on data extraction and treatment of physiological data recorded for this study, with the objective to develop a recognition system of stress based on machine learning. Indeed, this SSI offers the possibility to measure physiological data during stressful situation. Moreover, the capacities of SSI make it possible to modify the method of stimulating stress by imposing time constraints or cognitive tasks. The results presented were obtained by imposing time constraints; the next step will be to induce stress by imposing cognitive tasks on the individuals with the aim of exploring eventual specific physiological patterns. Another use of this application could be as a tool for stress sensor benchmarking.

References

1. Jones, F., Bright, J., Clow, A.: Stress: myth, theory, and research. Pearson Education (2001).
2. Van Der Vijgh, B., Beun, R.J., Van Rood, M., Werkhoven, P.: GASICA: Generic automated stress induction and control application design of an application for controlling the stress state. Frontiers in Neuroscience. 8, 1–17 (2014).
3. Stokes, A.F., Kite, K.: On grasping a nettle and becoming emotional. In: Stress, workload, and fatigue. pp. 107–132 (2001).
4. Tepas, D.I., Price, J.M.: What is stress and what is fatigue? In: Hancock, P.A. and Desmond, P. a (eds.) Stress, workload, and fatigue. Mahwah, N.J. : Lawrence Erlbaum Associates, 2001 (2001).
5. Selye, H.: The evolutioin of the stress concept. American Scientist. 61, 692–699 (1973).
6. Ivancevich, J.M., Matteson, M.T.: Stress and work : a managerial perspective. Scott Foresman, Glenview [etc.] (1980).
7. Lazarus, R.S.: Stress and emotion: A new synthesis. Springer Publishing Company. (2006).
8. Pickering, T.G., Devereux, R.B., James, G.D., Gerin, W., Landsbergis, P., Schnall, P.L., Schwartz, J.E.: Environmental influences on blood pressure and the role of job strain. J Hypertens Suppl. 14, S179–185 (1996).
9. Shi, Y., Ruiz, N., Taib, R., Choi, E., Chen, F.: Galvanic Skin Response (GSR) As an Index of Cognitive Load. In: CHI '07 Extended Abstracts on Human Factors in Computing Systems. pp. 2651–2656. ACM, New York, NY, USA (2007).
10. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems. 6, 156–166 (2005).
11. de Santos Sierra, A., Avila, C.S., Bailador del Pozo, G., Guerra Casanova, J.: Stress detection by means of stress physiological template. In: 2011 Third World Congress on Nature and Biologically Inspired Computing (NaBIC). pp. 131–136 (2011).
12. Partala, T., Surakka, V.: Pupil size variation as an indication of affective processing. International Journal of Human-Computer Studies. 59, 185–198 (2003).
13. Rubio, S., Diaz, E., Martin, J., Puente, J.M.: Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. Applied Psychology. 53, 61–86 (2004).
14. Dedovic, K., Renwick, R., Mahani, N.K., Engert, V., Lupien, S.J., Pruessner, J.C.: The Montreal Imaging Stress Task: Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. In: Journal of Psychiatry and Neuroscience. pp. 319–325 (2005).
15. Kirschbaum, C., Pirke, K.M., Hellhammer, D.H.: The "Trier Social Stress Test"--a tool for investigating psychobiological stress responses in a laboratory setting, (1993).

16. Ogden, G.D., Levine, J.M., Eisner, E.J.: Measurement of Workload by Secondary Tasks. Human Factors. 21, 529–548 (1979).
17. Campbell, J.I.D., Austin, S.: Effects of response time deadlines on adults' strategy choices for simple addition. Memory cognition. 30, 988–994 (2002).
18. Helton, W.S., Naswall, K.: Short stress state questionnaire: Factor structure and state change assessment. European Journal of Psychological Assessment. 31, 20–30 (2015).
19. Matthews, G., Campbell, S.E., Falconer, S., Joyner, L. a, Huggins, J., Gilliland, K., Grier, R., Warm, J.S.: Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. Emotion. 2, 315–340 (2002).
20. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A. and Meshkati, N. (eds.) Human Mental Workload. pp. 139–183. North-Holland (1988).
21. Byers, J.C., Bittner, A.C., Hill, S.G.: Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In: Advances in industrial ergonomics & safety. pp. 481–485 (1989).
22. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 67, 1–48 (2015).
23. Gueorguieva, R., Krystal, J.H.: Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch. Gen. Psychiatry. 61, 310–317 (2004).
24. Baayen, R.H., Davidson, D.J., Bates, D.M.: Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 59, 390–412 (2008).

Group effects on individual foraging in shoals of Zebrafish

Roy Harpaz and Schneidman Elad

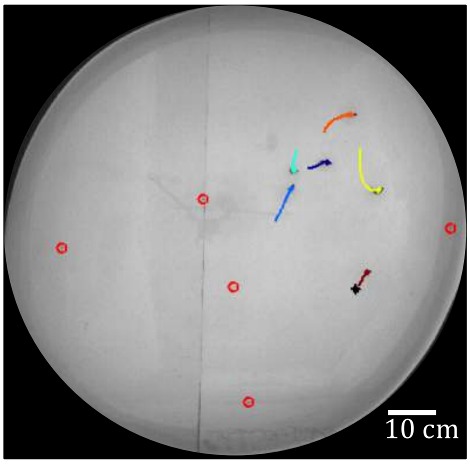
Weizmann Institute of Science, Israel

Collective foraging is often considered one of the prominent advantages of living together in a group and was studied extensively in different species form insects to humans [1]. Zebrafish (*Danio rerio*) live in small groups in nature (3-12 fish) [2], and while much is known about their genetics and individual behavior, our understanding of group behavior is lacking. Specifically, little is known about group foraging, and how information is transferred between shoal mates [3][4].

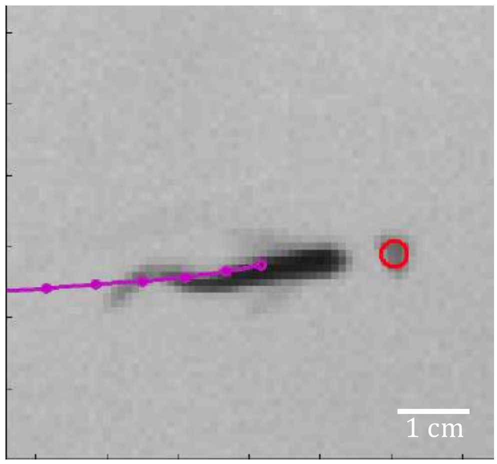
Here, we study the nature of foraging by groups of zebrafish in a laboratory environment and its dependence on group size, food distribution and individual differences. We propose a quantitative model of individual foraging of a fish in a group, which takes into account its response to the cues and the behavior of other fish.

Instead of using a structured 'forced choice' setup, in which animals choose between a finite and well determined feeders or food patches [5], [6], we studied foraging behavior of fish in an open arena, where food is scarce and difficult to detect, as is the case in their natural habitat (see Figure 1). We used groups of 1,3, and 6 mature zebrafish (>3 month), housed together in their designated groups for one month prior to the foraging experiments. All groups were trained to seek food in a larger arena (~ 95 cm in diameter) where approximately 6,12, or 18 small flakes of food were randomly scattered on the water surface. Fish swimming behavior were recorded at 50 fps before and after training. Individual fish were tracked using software written in Matlab, and fish identities were later corrected using IdTracker [7]. Our high-resolution videos allowed us to track and analyze individual and group foraging behaviors, specifically during food detections (Figure 1 B-D), and to build model simulations to test different hypothesized foraging strategies.

A.



B.



A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

C. D.

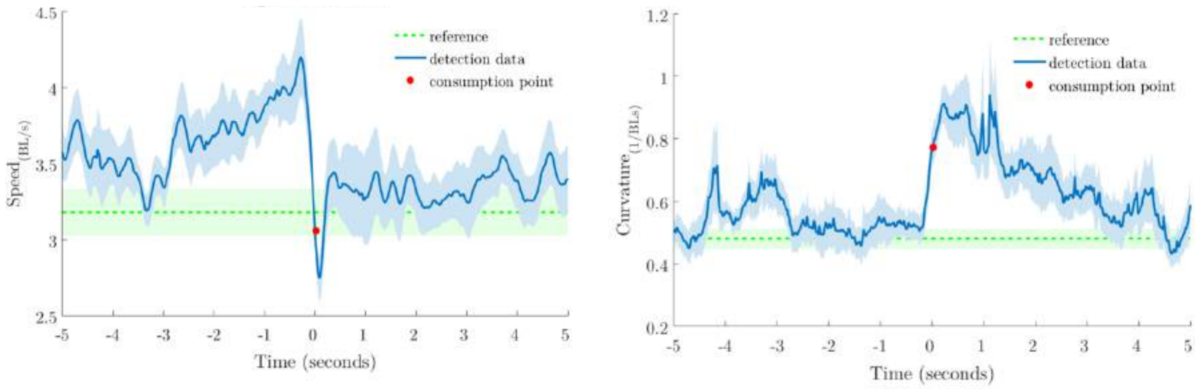


Figure 1: Free foraging in groups of zebrafish. **(A)** Experimental setup – groups of 1,3 and 6 fish were trained to search for small flakes of food (~4 mm in diameter) scattered randomly in shallow water (~5 cm) in a large circular arena (~95 cm in diameter), at a temperature of 24℃, and lit using 4 overhead incandescent light sources, 75W each. Colored lines designate individual fish trajectories, red circles are flakes of food, black asterisk are flakes previously consumed by fish. **(B)** Our high-resolution camera allow for automatic detection and tracking of fish trajectories, flakes of food and instances of flake consumptions. **(C-D)** fish show stereotypic behaviors around flake consumptions characterized by a transient decrease in speed and a typical turning (increase in trajectory curvature). Figures show mean properties over all detection events and all groups of 3 fish (blue lines), and a reference value calculated from random points along the trajectories not related to consumption events (green line).

Our results demonstrate that larger groups show an increased rate of flake consumption compared to single individuals (Figure 2A,B). To determine if this improvement is due to social interactions between fish or is it just due to the increase in the sheer number of independent searchers, we analyzed fish behavior at consumption events and tested these different possibilities using model simulations. Our behavioral analysis showed that individual fish performed specific and stereotypic behaviors while detecting and consuming food (figure 1C,D). Importantly, we found that neighboring fish were almost twice as likely to swim towards areas of previous food detection by other fish. We therefore hypothesized that the stereotypic maneuvers might serve as a signal for food availability.

A. B.

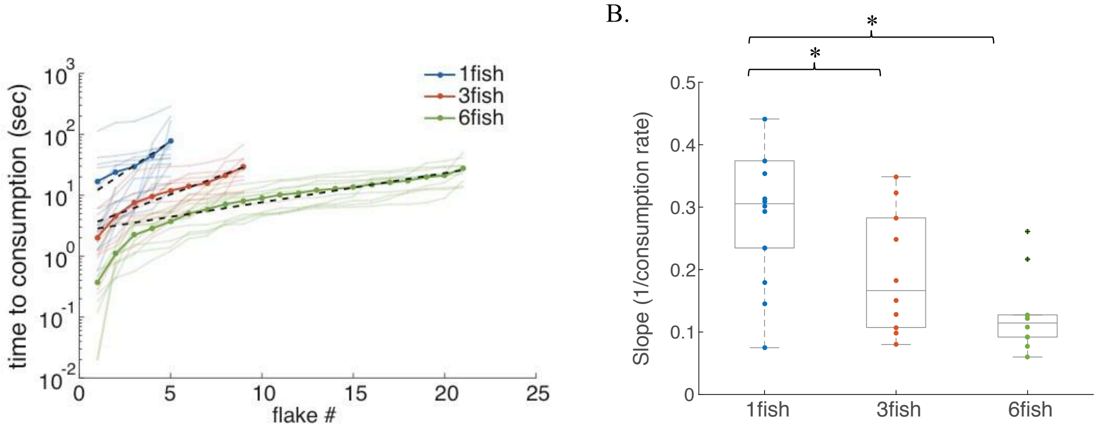


Figure 2: Increased rate of food consumption by zebrafish in a group. **(A)** The time it took the fish to consume the flakes in the arena (i.e. time to first flake, time to second flake, etc.) is shown for each of the groups tested (soft lines, 10 repetitions for each group size) overlaid with the mean of all groups of each size (bold lines). An exponential model of the form $T = a \cdot e^{bx}$ (dashed black line) closely matches the mean data (model parameters for 1 fish: $b = 0.367(0.25\ 0.48), a = 11.12(7.6\ 16.25)$, 3 fish: $b = 0.287(0.21\ 0.36), a = 2.4(1.59\ 3.64)$, 6 fish: $b = 0.156(0.13\ 0.19), a = 1.36(0.87\ 2)$ 95%

confidence interval in parenthesis). Note that the rate of flake consumption is higher for larger groups (i.e. the slope of the fitted line is smaller). **(B)** Boxplot showing the slopes of the same model $T = a \cdot e^{bx}$ fitted individually for each of the groups shown in (A). Horizontal lines represent median values; box edges are the 25th and 75th percentiles, asterisk represent statistically significant differences (P<0.05). Again, larger groups are faster at finding flakes than single fish.

To test this potential attraction mechanism, we used model simulations based on fish swimming properties (Figure 3). Such a modeling scheme can also allow us to test whether a specific social interaction mechanism has an advantage over a simple group of independent foragers.

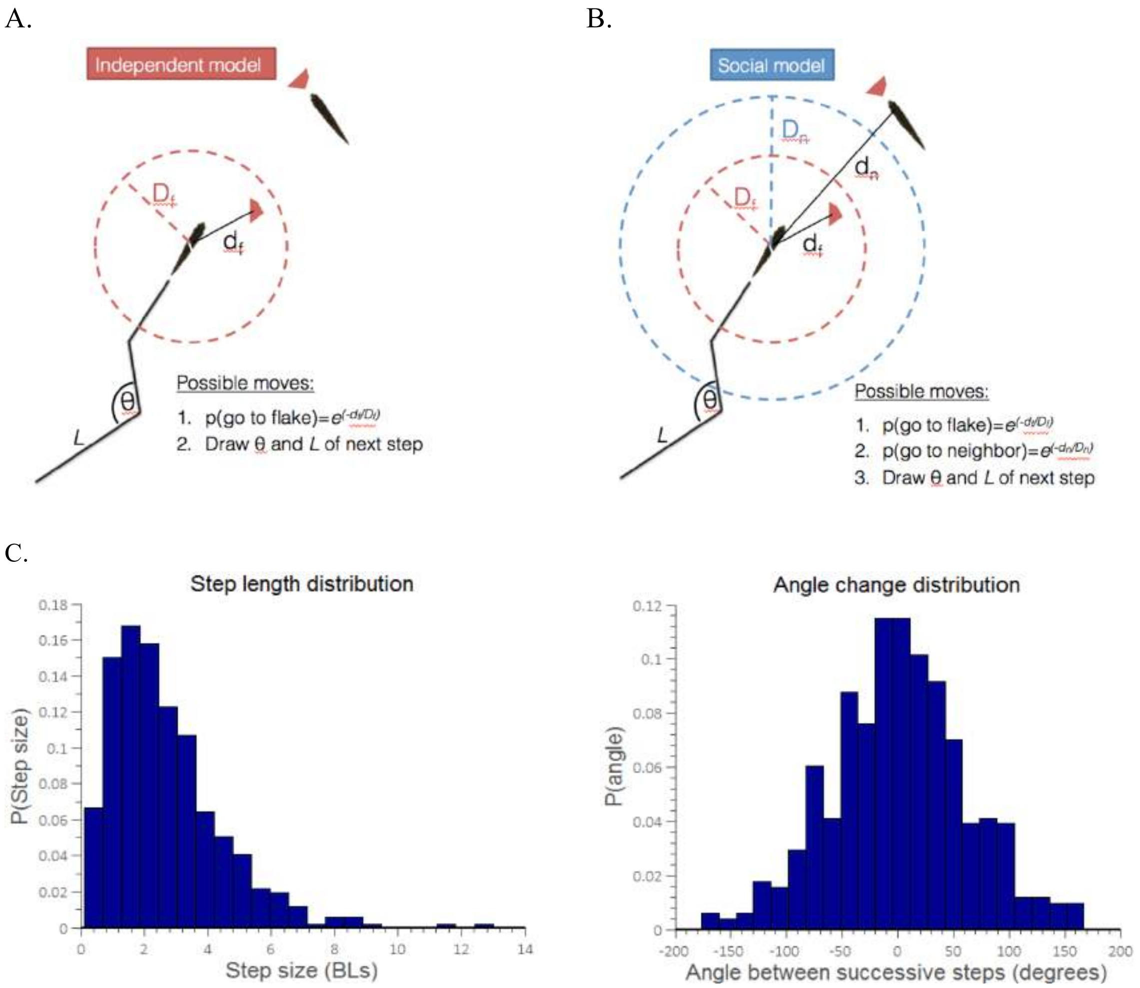


Figure 3: Testing fish foraging strategies using independent and social models. To test our hypothesized social foraging mechanism we used three models of fish foraging. **(A)** An independent model of fish foraging (IND) in which at each time step we check if a flake is present within a fish's detection range ($d_f < D_f$ see red circle in figure) which will then cause the fish to go towards that flake with a probability p(go to flake). Otherwise the fish will make a move of length L and an angle change θ, taken from distributions estimated from real fish swimming (see step length and angle change distributions in panel (C)). **(B)** For comparison, we used 2 social models: one based on social attraction (ATT) and one based on social repulsion (REP), in which in addition to flake detection, fish can also detect neighbors consuming flakes ($d_n < D_n$ see blue circle in figure). If such a detection was made, fish will attract or repel form that point with probability p(go to/away from neighbor) in accordance with the social model used. (**C**) Distribution of step size L and angle change θ, between discrete steps. To obtain these distribution fish trajectories were discretized into 'steps' using the minima points of fish speed profile (data not shown here).

We found that an attraction-based model indeed gave the best fit to our data and outperformed other possible strategies, namely independent foraging or repulsion from detection areas (Figure 4 A-C). Moreover, a comprehensive exploration of the parameter space of our models showed that simulated groups using an attraction strategy would be more efficient than simulated independent foragers only in a specific part of the parameter space. We found that all real foraging groups were best matched by simulated groups with parameters that were in this range, which supports our hypothesis that groups increase their efficiency due to social interactions (Figure 4D). Finally, we found only a weak relationship between the density of food resources in space (how much food was clustered together in one location), and the strategy used by our fish, or a consistent relationship between individual properties and food intake (data not shown).

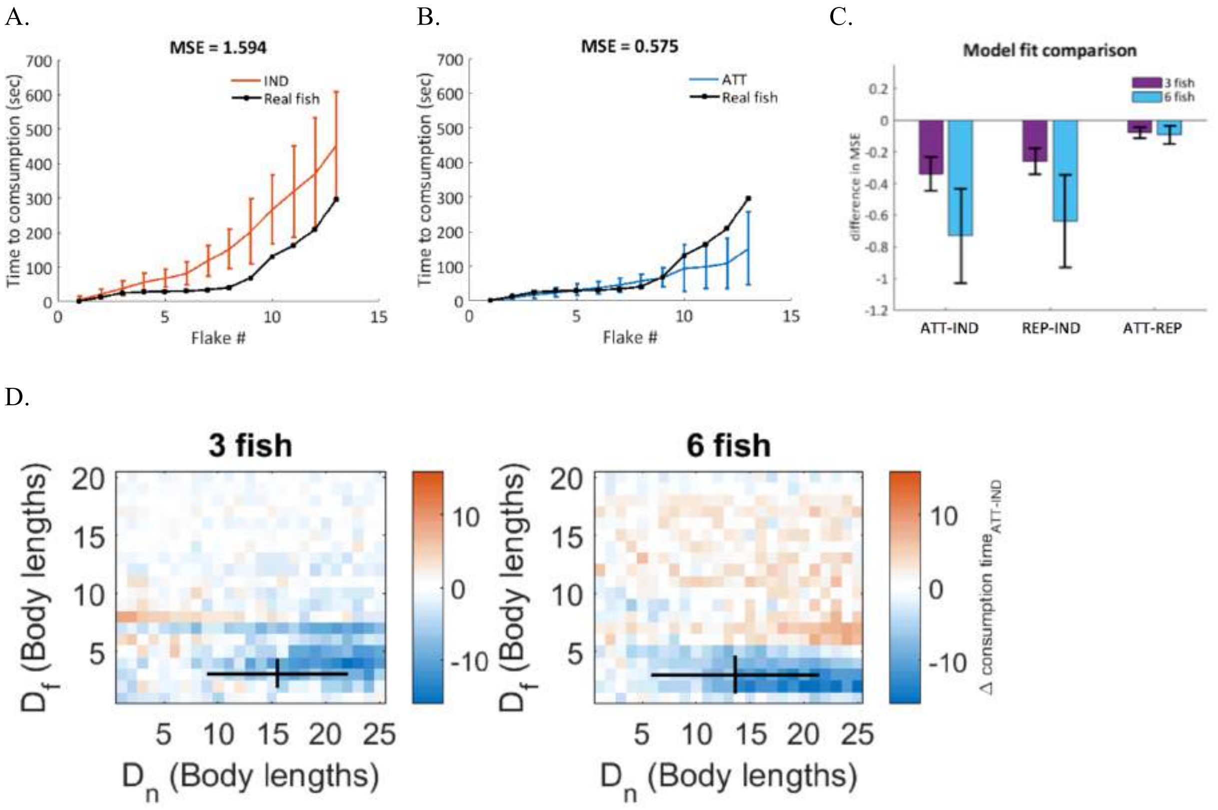


Figure 4: Attraction based models show the best fit to real foraging groups. **(A)** Example of flake detection times of a 3 fish group and the average and standard deviation of the best-fit independent foraging model (red). B) The same group as in (A) with the average and standard deviation of the best-fit social attraction model (blue). Notice the improvement in the ability of the model to describe flake detection times. **(C)** Comparison of mean squared error (MSE) between real and predicted detection times for all models and for different group sizes. Incorporating an attraction term significantly improves the fit of the models, reducing the MSE by an average factor of 2.44 ± 1.6 compared to the IND model. Using a repulsion term also significantly improves the fit by an average factor of 1.83 ± 0.64. The addition of an attraction term is significantly more accurate than a repulsion term, although we cannot rule out a combination of the two strategies in different groups. **(D)** Comparison of consumption times of simulated groups using an attraction strategy to those using an independent strategy for all parameter combinations. Blue areas in our figures are parameter regimes in which the ATT models improves searching efficiency while red areas are regimes in which social foraging actually hinders foraging. Black crosses represent the range of parameters of simulated groups that best fitted our real fish groups, which are well in the areas where social foraging is beneficial in terms of foraging efficiency.

Understanding how individuals in a group interact to increase foraging capabilities is an important challenge in the field of collective animal behavior. Our results show for the first time how interactions between individuals can increase foraging capabilities of zebrafish groups. The exact nature of information transfer between fish, and the modalities in which information flows between individuals – visual, olfactory, tactile or some combination of them – calls for additional studies.

All experimental procedures were reviewed and authorized by the IACUC committee of the Weizmann Institute of Science.

References

1. L.-A. Giraldeau and T. Caraco, Social Foraging Theory. Princeton University Press, 2000.
2. R. Spence, G. Gerlach, C. Lawrence, and C. Smith, "The behaviour and ecology of the zebrafish, *Danio rerio*," Biological Reviews, vol. 83, no. 1, pp. 13–34, Feb. 2008.
3. J. Krause, N. Hartmann, and V. L. Pritchard, "The influence of nutritional state on shoal choice in zebrafish, *Danio rerio*," Animal Behaviour, vol. 57, no. 4, pp. 771–775, Apr. 1999.
4. J. R. G. Dyer, D. P. Croft, L. J. Morrell, and J. Krause, "Shoal composition determines foraging success in the guppy," Behavioral Ecology, vol. 20, no. 1, pp. 165–171, Jan. 2009.
5. N. Miller, S. Garnier, A. T. Hartnett, and I. D. Couzin, "Both information and social cohesion determine collective decisions in animal groups," PNAS, Feb. 2013.
6. A. J. W. Ward, J. Krause, and D. J. T. Sumpter, "Quorum Decision-Making in Foraging Fish Shoals," PLoS ONE, vol. 7, no. 3, p. e32411, Mar. 2012.
7. A. Pérez-Escudero, J. Vicente-Page, R. C. Hinz, S. Arganda, and G. G. de Polavieja, "idTracker: tracking individuals in a group by automatic identification of unmarked animals," Nat Meth, vol. 11, no. 7, pp. 743–748, Jul. 2014.

Symposia

Temporal and Spectrographic complexity and structure in mouse USVs: The Problem of Meaning

Katrin Schenk

Randolph College, United States of America

Many neuropsychiatric diseases are associated with communication and/or social deficits. Mouse Ultrasonic Vocalizations have been used by many investigators as an assay for these deficits in mouse models of disease. However, all of these assays suffer from the same fundamental problem, the lack of a connection between the spectral and temporal complexity of mouse sultrasonic calls and behavioral and emotional significance. For instance, in our 2010 PNAS paper [Young et al, 2010] we analyzed the ultrasonic vocalizations (USVs) of mice with a genetic mutation that causes Tuberous Sclerosis, a disease highly associated with autism. Our assay consisted of quantifying several spectral and temporal quantities of the calls as well as classifying the calls into several categories according to their visual morphology on a spectrogram. Although we found differences in several parameters including call type between the mutant and wild-type mice, these differences may not have represented real differences in social or communication abilities. For example, if mutant mice are found to make more of a certain type of call than wild-type mice it may be because these overrepresented calls signify a different emotional state than other calls. If this were true, then this difference would not be attributable to any high level social or communication processing or production deficit but would simply represent a propensity for a different emotional state. To gain a clearer understanding of any differences one finds between mouse models of disease and their wild-type counterparts one must understand the emotional context and behavioral meaning of mouse USVs.

We will discuss several problems centered around the interpretation of mouse USVs and present several approaches that may prove fruitful in addressing these issues.

Reference

Young, D., Schenk, A., Yang, S.-B., Jan, Y. & Jan, L. Altered ultrasonic vocalizations in a tuberous sclerosis mouse model of autism. Proceedings of the National Academy of Sciences of the United States of America **107**, 11074–9 (2010).

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Elucidating Complex, Innate Mouse Social Behavior

J.P. Neunuebel

Dept. of Psychological and Brain Sciences, University of Delaware, Newark, Delaware. USA.
jneun@udel.edu

During courtship males perform elaborate behaviors to attract females. In mice, these innate behaviors include ultrasonic vocalizations. Although ultrasonic courtship vocalizations were previously attributed to the male, evidence exists that both sexes produce indistinguishable vocalizations. Because of this similarity, and the challenge of assigning vocalizations to individuals during group interactions, it is unknown how each animal vocally contributes to courtship. To address this question, we developed a microphone array system to localize vocalizations from socially interacting, individual adult mice. This system allowed us to discover that female mice vocally interact with males during courtship. More specifically, males and females jointly increased their vocalization rates during courtship chases. Moreover, a female's participation in these vocal interactions may function as a signal indicating a state of increased receptivity. Our results reveal a novel form of vocal communication during mouse courtship, and lay the groundwork for a mechanistic dissection of communication during social behavior.

All experiments were performed at Janelia Research Campus in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The Janelia Research Campus Institutional Animal Care and Use Committee (IACUC; protocol #11-70) approved every procedure that was carried out in the described studies. The Janelia Research Campus Vivarium maintains full AAALAC accreditation.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Elucidating USVs; vocalizing and receiving

Hye Young Lee[1], Shih-Yi Tseng[2], Kunwoo Lee[3], Katrin Schenk[4]

[1]The University of Texas, Health Science Center at San Antonio, United States of America
[2]National Taiwan Univeristy, Taiwan
[3]University of California, Berkeley, United States
[4]Randolph college, United States

Autism spectrum disorder (ASD) and autism are both general terms for a group of complex disorders of brain development. These disorders are characterized, in varying degrees, by difficulties in social interaction, verbal and nonverbal communication and repetitive behaviors. Today, one out of every 68 children is affected with autism or a related disorder and an estimated over 3 million individuals in the U.S. and tens of millions worldwide are affected by autism. Thus, it is now more common than childhood cancer, juvenile diabetes and pediatric AIDS combined. About half of all individuals with autism fail to develop fully functional speech, and for those who do, language is generally characterized by one or more abnormalities such as pronoun reversal, use of neologisms, stereotyped or rigid speech, and abnormalities in intonation. A history of language delay can be particularly crucial in differentiating autism from other psychiatric disorders in high-functioning adults. Although this language delay is one of major symptoms of autism and furthermore, milestones in language and communication play major roles at almost every point in development in understanding autism, neurobiology underlying this social communication deficit has not been studied enough to explain what the possible causes and the mechanisms are in autism, mainly because of the lack of animal models and measuring system.

Although it is difficult to recapitulate the language disorders in animal models, ultrasonic vocalization (USV) is a consistent and robust phenomenon in rodents during adult social interactions and can be considered as an index of social interest and motivation and may be a measure of social communication in mice. Moreover, adult mouse vocalizations have been positively correlated with social investigation in social interactions. Our preliminary results show abnormal social interactions and social communications in adult male fmr1 KO mouse, a mouse model of one of a few single gene-driven autism Fragile X Syndrome (FXS), when interacting with female WT mouse. Interestingly, fmr1 KO males chase WT females normally, but the mutual interaction is significantly decreased in fmr1 KO male with WT female, which correlates with the reduction of USV call numbers and the increase of latency to the first call observed in fmr1 KO males. We hypothesize that females respond differently to WT and fmr1 KO male USVs, which mediate their mutual interactions reduced when interacting with fmr1 KO males. Here, my talk will focus on uses of USVs as a behavioral measurement in fmr1 KO mice with these two aspects; vocalizing and receiving.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Context-related variations in mouse ultrasonic vocalisations

E. Ey[1], A-T. Ferhat[1], A-M. Le Sourd[1], F. de Chaumont[2], J-C. Olivo-Marin[2] and T. Bourgeron[1]

1Human Genetics and Cognitive Functions, CNRS UMR 3571, University Paris Diderot, Sorbonne Paris Cité, Institut Pasteur, Paris, France.
elodie.ey@pasteur.fr; aferhat@pasteur.fr; lesourd@pasteur.fr; thomas.bourgeron@pasteur.fr
2Bio Image Analysis, CNRS URA 2582, Institut Pasteur, Paris, France.
fabrice.de-chaumont@pasteur.fr; jcolivo@pasteur.fr

Rodent ultrasonic vocalisations are now broadly used as proxies for communicative abilities in behavioural characterisations of models for neuropsychiatric diseases [1,2]. Test conditions vary between different models. The effect of these test conditions on the quantity and quality of social interactions and ultrasonic communication is rarely exploited.

Influence of test cage and habituation time on mouse ultrasonic vocalisations

We examine to what extent the habituation time as well as the shape / size of the test cage influence social communication in freely interacting male mice. Adult mice are recorded in a test cage with fresh bedding. After a habituation period, we introduced an unknown male mouse and leave the two animals interact [3]. We tested two habituation times (20 min and 30 min). We also compared the social interactions and ultrasonic vocalisations emitted in three different types of cage (rectangle, square and round). We did not find major effects of habituation time and test cage style, suggesting that a short habituation time within a small rectangular cage is sufficient not to force social interactions [4].

Influence of previous social experience on mouse ultrasonic vocalisations

It is known that adult males need to be socially isolated, so that aggressive behaviours are diminished and ultrasonic vocalisations are recorded. This is much less obvious in female-female interactions. Nevertheless, we observed that a short previous social isolation increase the motivation of the tested mice. More ultrasonic vocalisations are emitted and more social contacts are done in such cases (unpublished data).

New insights from spontaneous social communication in mice

Current experimental settings do not allow mice to express their complete behavioural repertoire. Indeed, the motivations of the animals are manipulated and we constrain the social interactions to dyadic interactions. The artificiality of experimental conditions does not allow the animals to express their complete natural behavioural repertoire and might lead to difficulties in reproducing results. We therefore suggest developing innovative test settings, allowing the animals to be tested while they express a larger part of their spontaneous behavioural repertoire. We present a preliminary setting to record spontaneous social interactions and ultrasonic vocalisations continuously over several days and nights (unpublished data).

All data collected are made available to the whole community on the *mouseTube* database (http://mousetube.pasteur.fr) to boost knowledge on mouse ultrasonic communication. Such approaches will deepen our understanding of rodents as such, and therefore lead to more reliable characterisation of behavioural phenotypes in rodent models of neuro-psychiatric diseases. Procedures involving animal subjects have been approved by the Comité d’Ethique en Expérimentation animale (CETEA) n°89 at the Institut Pasteur, Paris.

References

1. Scattoni, M.L., Crawley, J., Ricceri, L. (2009) Ultrasonic vocalizations: A tool for behavioural phenotyping of mouse models of neurodevelopmental disorders. Neuroscience Biobehavioral Reviews 33: 508–515.

doi:10.1016/j.neubiorev.2008.08.003.
2. Silverman, J.L., Yang, M., Lord, C., Crawley, J.N. (2010) Behavioural phenotyping assays for mouse models of autism. Nature Reviews Neuroscience 11: 490–502. doi:10.1038/nrn2851.
3. Ferhat, A-T., Torquet, N., Le Sourd, A-M., de Chaumont, F., Olivo-Marin, J-C., Faure, P., Bourgeron, T., Ey, E. (in press) Recording mouse ultrasonic vocalizations to evaluate social communication. Journal of Visualized Experiments.
4. Ferhat, A-T., Le Sourd, A-M., de Chaumont, F, Olivo-Marin, J-C., Bourgeron, T., Ey, E.. (2015) Social communication in mice – Are there optimal cage conditions? PLoS One 10(3), DOI: 10.1371/journal.pone.0121802.

Overview of Key Issues Associated With Using Nocturnal Rodents during the Working Day

P. Hawkins[1] and H. Golledge[2]

[1]Royal Society for the Prevention of Cruelty to Animals (RSPCA), United Kingdom
penny.hawkins@rspca.org.uk
[2]Universities Federation for Animal Welfare UFAW), United Kingdom
golledge@ufaw.org.uk

Introduction

Behavioural studies (and many others) using rats and mice typically aim to understand the 'normal' behavioural responses of these animals, or the effects of disease 'models' or treatments on these responses. However, studies are usually conducted in the light phase during the human working day. Since mice and rats are nocturnal, or crepuscular, this seems like an unnatural time to test them – when they would normally be resting or sleeping [1]. How does this affect both the quality of data obtained and the welfare of the animals?

Scientific implications

Testing during the human daytime represents a manifestly unnatural situation if one wishes to study the 'normal' cognitive or behavioural capacity of a nocturnal animal, since behaviour and cognition can differ significantly when nocturnal animals are tested during the light and dark phases. For example, DBA mice are thought to be an animal model of anxiety and have also been reported to exhibit cognitive dysfunction in spatial memory tests. However, behavioural inhibition, an indicator of anxiety, was significantly reduced when mice were tested using a modified hole board during the dark (or active) phase as opposed to the light phase. The cognitive performance of the mice was also better at times when they would naturally be awake and active [2].

Differences in physiological responses have also been noted; for example, a significant increase in the 50% withdrawal threshold to von Frey hair stimulation (i.e. pain threshold) has been reported in C57BL/6 mice during the light phase as opposed to the dark, active phase, which has implications for testing analgesics and may result in a pain-related behavioural phenotype being misinterpreted or missed [3]. Significant day-night differences in behaviour, bladder capacity and micturition frequency have been recorded in rats [4], which could affect the translatability of research aiming to address bladder dysfunction in humans.

Conversely, others have found that the light-dark cycle did not have any discernible effects on the responses of mice, following the administration of ethanol, on balance beam, grip strength or accelerating rotarod tests [5]. The authors discuss reasons for this, including the potential for noises within the facility that could not be controlled (such as elevators) acting as zeitgebers, or the 30 minute acclimation period before testing possibly arousing the mice. Another study reported no main effects of circadian phase on outcome in behavioural tests such as open field, elevated plus and water mazes and hyperactivity in response to amphetamine, similarly suggesting that 'being tossed in a large, cold pool of water (or poked with a needle, or placed in narrow lanes with a precipitous drop in an unfamiliar place) may simply engage a response system not subject to significant circadian fluctuation' [1]. These authors speculate as to whether behavioural tests that target more refined behavioural phenotypes, and/or are less stimulating or disruptive, may be more affected by circadian phase.

The take-home message, from a scientific viewpoint, is that there can be significant variation in responses depending on the light-dark cycle and time of day when animals are tested, and this may give misleading results. This will have implications for both basic research and the drug discovery pipeline, and therefore for patient groups, as well as creating ethical issues because animals will have been wasted. Another potential consequence, also with ethical and animal welfare implications, would be increased variability in data causing additional animals to be required to achieve a statistically significant result.

A similar conundrum has been faced in visual neuroscience, where unnatural (over-simplified) stimuli have limited the understanding of the response of the brain to visual input, but subsequent use of complex natural images has allowed a more detailed understanding of the response of the system [6]. For example, the responses of the cat visual cortex differ qualitatively and quantitatively when presented with natural scenes as opposed to artificial stimuli [7]. Such results strongly argue that, wherever possible, biological systems should be studied whilst performing the functions for which they have evolved [8]. This is not limited to visual responses, and a case can be made that studying behaviour during the waking day will generally reveal more useful information about the 'normal' response of the animals and systems under study.

Further reflection on the biology and behaviour of rats and mice, in contrast to the way in which they are housed in the laboratory, shows that there many conflicts between the needs of humans and rodents (Table 1) [9]. Adopting a naturalistic approach to rodent behaviour testing, with respect to circadian rhythms at least, is an approach which could reduce one such conflict and improve data quality and translatability.

Table 1. Conflicts between the needs of humans and laboratory mice and rats

Mice and rats are ...	... but
Nocturnal and crepuscular	they are housed in bright light and used during the day when they would normally be sleeping
Highly dependent on smell and scent markings	their markings are completely destroyed whenever the cage is cleaned
Sensitive to ultrasound	there are many sources of ultrasound in the laboratory, and these are not always checked and minimised
Able to feel more secure when touching objects (thigmotaxis)	they are often housed in barren cages
Master diggers	they have no opportunity to burrow
Highly social (species, sex and strain dependent)	they are often housed in inappropriate groups or singly
Capable of covering long distances	they are housed in small cages
Omnivorous, trying new foods from different feeding sites	they are fed boring, monotonous diets from hoppers
Made extremely anxious when captured by the tail (mice)	most people catch them by the base of the tail
Most comfortable at a temperature of 26 to 34 ^{o}C (mice)	many facilities house them at colder temperatures, and/or do not provide sufficient nesting material

On the basis of the above, it should never be assumed that light-dark cycles and time of testing are irrelevant. There is a case for taking a standard approach of systematically testing nocturnal animals during their active period – unless there is a specific need to measure responses during the light (inactive) phase.

Animal welfare implications

In addition to the potential negative impacts on the science outlined above, there are animal welfare implications associated with disturbing animals when they would normally be asleep. Mice sleep deeply at the beginning of the light phase, which is often corresponds to morning cage change or procedures. It has been suggested that time-shifting mice so that they are deeply asleep in the early hours of the morning, and ready for activity at the beginning of the human working day, could therefore make a significant improvement to the quality of the life of the laboratory mouse [10].

Evidence in the literature supports this; for example, disruption of the light cycle causes increased aggression and indicators of stress amongst mice [11]. Rodents also seem more prone to be affected by stressors when they occur during their resting period, in that stressors which induce anxiety or depression-like symptoms when applied during the light phase do not cause such symptoms when they occur during the dark [12]. Even a test specifically designed to cause stress, such as the forced-swim test, causes a smaller stress response when animals are tested in the dark [13]. This strongly suggests that, where testing procedures are inherently stressful (the Morris water maze, for instance) the welfare of the animals will be further compromised by testing them during the light phase, so this should not be undertaken unless there is a justifiable scientific requirement to elicit a more extreme response.

One approach to avoiding these welfare problems and scientific confounds is to test rodents during actual or simulated night-time, usually by maintaining a reversed light cycle, whereby the room in which procedures are conducted is maintained in darkness during the human working day. This approach is appearing more frequently in the literature [14], but is not yet the norm. Mice will entrain to low amplitude dim/bright light cycles, noting that there are strain differences, and it has been suggested that this could be a strategy for standardising and managing circadian phase in nocturnal animals [1].

Taking steps to reduce, or eliminate, avoidable stress to animals by conducting procedures when they would normally be awake will therefore reduce experimental harms, improve welfare and improve compliance with legislation regulating animal use. For example, European Commission Directive 2010/63/EU requires the refinement of methods used in procedures, minimising distress, and mandates that ‘any restrictions on the extent to which an animal can satisfy its physiological and ethological needs are kept to a minimum’.

Practical factors

How easy is it to maintain animals on a reversed light-cycle? Two obvious issues are the need to completely avoid any flashes of light, however brief, which could disrupt the animals’ circadian rhythm, and the need for researchers and caregivers to be able to see the animals in order to care for and monitor them, and to obtain data.

Night vision goggles are one option [5], or reversing the working day for humans, or the use of red lights and red film covering viewing panels is more commonplace. An alternative approach is the use of lighting which allows adequate vision for researchers whilst containing wavelengths at the extremes of rodent visual spectral sensitivity. For example, humans have good visual acuity in sodium light, whereas it falls on the very margins of the visual sensitivity of the mouse, such that mice appear to display normal nocturnal locomotory behaviour in sodium light [10]. Mice do have some sensitivity to red light, so using red lighting adequate for human vision may not ensure that mice perceive this as subjective night time. Whatever the lighting protocol, there will inevitably be implications for animal house management, but these should be justified by the positive impacts on both science and animal welfare. It would be helpful if researchers, animal technologists and animal unit managers could exchange information on good practice protocols for reversing light regimes.

Reporting issues

There is a clear and indisputable need to clearly define the lighting conditions as part of routine reporting of expriments, to raise awareness of extraneous experimental variables [15]. Both the Gold Standard Publication Checklist and the ARRIVE guidelines include the need to include information on lighting regimes and the time of day at which procedures were conducted [16,17], and the point has been made that circadian effects should be acknowledged and controlled for in the same way as strain and sex [1].

As an example of good practice in this regard, Munn et al. (2011) discuss the potential effects of light-dark reversal on their data, describe the light regime in detail and discuss why there was no apparent effect on the parameters they were testing [5].

Conclusions

When undertaking behavioural research using mice and rats during the light phase, the working assumption should be that there <u>will</u> be effects on data quality unless proven otherwise. Even if the behavioural tests in question appear to be unaffected by light regime and time of day – as demonstrated by pilot studies – the default should still be to conduct experiments during the active (dark) phase for reasons of animal welfare, unless there is a scientifically justifiable reason otherwise.

There may be objections to changing routine light regimes for laboratory mice and rats on the grounds of data compatibility, but the assumption should be that, if data are different, the data quality has been improved by taking a more 'naturalistic' approach. Any investment in the animal house will be justified by the enhancements to science and animal welfare.

One immediately achievable change in practice would be for all authors to include details of light regimes and timings of behavioural tests in all publications, with a discussion of the potential impact of these on data quality. This also depends upon journal editors supporting this level of information and including it in their requirements for authors. In the short term, this would help to achieve greater recognition and acknowledgement of the scientific and animal welfare implications of using nocturnal animals during the light phase, within the human working day.

References

1. Beeler, J.A., Prendergast, B., Zhuang, X. (2006). Low amplitude entrainment of mice and the impact of circadian phase on behavior tests. Physiology & Behavior 87, 870-880.
2. Roedel, A., Storch, C., Holsboer, F., Ohl, F. (2006). Effects of light or dark phase testing on behavioural and cognitive performance in DBA mice. Laboratory Animals 40, 371–381.
3. Minett, M.S., Eijkelkamp, N., Wood, J.N. (2014). Significant determinants of mouse pain behaviour. PloS ONE 9(8): e104458. doi:10.1371/journal.pone.0104458.
4. Herrera, G.M., Meredith, A.L. (2010). Diurnal variation in urodynamics of rat. PloS ONE 5(8): e12298. doi: 10.1371/journal.pone.0012298.
5. Munn, E., Bunning, M., Prada, S., Bohlen, M., Crabbe, J.C., Wahlsten, D. (2011). Reversed light-dark cycle and cage enrichment effects on ethanol-induced deficits in motor coordination assessed in inbred mouse strains with a compact battery of refined tests. Behavioral Brain Research 224, 259-271.
6. Felsen, G., Dan, Y. (2005). A natural approach to studying vision. Nature Neuroscience 8, 1643-6.
7. Kayser, C., Salazar, R.F., and Konig, P. (2003). Responses to natural scenes in cat V1. J. Neurophysiol. 90, 1910–1920.
8. Kayser, C., Körding, K.P., and König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. Curr. Opin. Neurobiol. 14, 468–473.
9. Liss, C., Litwak, K., Tilford, D., Reinhardt, V. (eds). (2015). Comfortable Quarters for Laboratory Animals, 10th edn. Animal Welfare Institute, Washington, D.C. <http://awionline.org/sites/default/files/products/AW I-ComfortableQuarters-2015.pdf>. Accessed 27 January 2016.
10. McLennan, I.S., Taylor-Jeffs, J. (2004). The use of sodium lamps to brightly illuminate mouse houses during their dark phases. Laboratory Animals 38, 384–392.
11. Meer, E.V. der, Loo, P.L.P.V., and Baumans, V. (2004). Short-term effects of a disturbed light–dark cycle and environmental enrichment on aggression and stress-related parameters in male mice. Laboratory Animals 38, 376–383.
12. Aslani, S., Harb, M.R., Costa, P.S., Almeida, O.F.X., Sousa, N., and Palha, J.A. (2014). Day and night: diurnal phase influences the response to chronic mild stress. Frontiers in Behavioral Neuroscience 8:82 doi: 10.3389/fnbeh.2014.00082.
13. Kelliher, P., Connor, T.J., Harkin, A., Sanchez, C., Kelly, J.P., and Leonard, B.E. (2000). Varying responses to the rat forced-swim test under diurnal and nocturnal conditions. Physiology & Behavior 69, 531–539.
14. Leger, M., Quidevile, A., Bouet, V., Haelewyn, B., Boulouard, M., Schumann-Bard, P., Freret, T. (2013). Object recognition test in mice. Nature Protocols 8, 2531-2537.
15. Prager, E.M., Bergstrom, H.C., Grunberg, N.E., Johnson, L.R. (2011). The importance of reporting housing and husbandry in rat research. Frontiers in Behavioral Neuroscience 5:38. 10.3389/fnbeh.2011.00038.
16. Hooijmans, C.R., Leenaars, M., Ritskes-Hoitinga, M. (2010). A Gold Standard Publication Checklist to Improve the Quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. ATLA 38, 167–182.
17. NC3Rs. (2010). <https://www.nc3rs.org.uk/arrive-guidelines>. Accessed 27 January 2016.

Light and the laboratory mouse

Stuart Neil Peirson

University of Oxford, United Kingdom

Introduction

Light exerts widespread effects on physiology and behaviour. As well as the widely-appreciated role of light in image-forming vision, light also plays a critical role in many non-image forming responses, including regulating circadian rhythms and acute light responses including modulating sleep induction, pupil constriction, heart rate, hormone release and learning and memory [1]. In mammals, these responses are all mediated via retinal photoreceptors, including the classical rods and cones involved in vision as well as the recently identified melanopsin-expressing photoreceptive retinal ganglion cells (pRGCs) [2]. Understanding the effects of light on the laboratory mouse therefore depends upon an appreciation of the physiology of these retinal photoreceptors, including their differing sensitivities to absolute light levels, as well as their responses to different wavelengths (colours). Moreover, the signals from these photoreceptors are often integrated, and different responses involve different central signalling pathways, making generalisations challenging [1].

Circadian regulation of behaviour

Several studies have assessed the impact of light and dark on behavioural testing [3]. In DBA mice, testing in the light has been shown to result in behavioural inhibition and cognitive disruption [4]. Conversely, testing in the dark results in improved discrimination in a range of behavioural tests, including the widely used SHIRPA test battery [5]. By contrast, there have been relatively few studies that have assessed the effects of circadian phase on performance in behavioural tests (that is, under constant conditions). In some cases, no effect of circadian phase was found on a range of behavioural tests [6]. However, other studies have demonstrated a notable impact of circadian phase on learning and memory, which would be expected to translate into performance. Surprisingly, mice acquired the conditioning faster in the day than in the night [7].

Circadian and photic effects on learning and memory, visual function and sleep

To test the effects of circadian phase on learning and memory, we conducted a counterbalanced, within-groups study of novel object recognition to compare performance during day and night, under both light/dark cycles and constant conditions. In agreement with previous data, we found that performance was consistently better during the subjective day. In a separate study, we have also assessed the effects on light/dark and constant conditions on visual function in mice. These studies show clear reductions in photopic ERG, contrast sensitivity and pupillary light responses during the subjective night. Finally, studies on nocturnal light exposure show that whilst such conditions can induce sleep in mice [8], they can also produce an arousal response associated with impaired sleep induction and elevated plasma corticosterone. These opposing responses are dependent upon the incident wavelength of light and appear to involve different neural pathways.

Practical considerations

Previous studies have suggested that reversing the lighting conditions in animal facilities and using sodium lamps during the day may provide a way of testing animals at a more biologically appropriate time [9]. This is based on the fact that mice have a reduced sensitivity to longer wavelengths. However, mice are only 8 times less sensitive to 590nm sodium lighting than humans. Data on action spectra for various biological responses in mice have shown that the threshold sensitivity of these responses is very low [10]. As such, the level of 'nocturnal' light for humans to operate would almost certainly produce biological responses in mice. This would effectively result in exposure to dim light at night – conditions that are known to produce a wide range of

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

adverse effects [11]. In addition to these issues, given the impairments in cognitive performance and visual function during the night described above, reversing animal facility lighting may not be a simple solution to the issue of testing at a biologically relevant time of day.

Ethical Statement

All experiments were conducted in accordance with the Animals (Scientific Procedures) Act 1986 and the University of Oxford Policy on the Use of Animals in Scientific Research.

References

1. Lucas, R.J., et al., *Measuring and using light in the melanopsin age.* Trends Neurosci, 2014. **37**(1): p. 1-9.
2. Hankins, M.W., S.N. Peirson, and R.G. Foster, *Melanopsin: an exciting photopigment.* Trends Neurosci, 2008. **31**(1): p. 27-36.
3. Peirson, S.N. and R.G. Foster, *Bad light stops play.* EMBO Rep, 2011. **12**(5): p. 380.
4. Roedel, A., et al., *Effects of light or dark phase testing on behavioural and cognitive performance in DBA mice.* Lab Anim, 2006. **40**(4): p. 371-81.
5. Hossain, S.M., B.K. Wong, and E.M. Simpson, *The dark phase improves genetic discrimination for some high throughput mouse behavioral phenotyping.* Genes Brain Behav, 2004. **3**(3): p. 167-77.
6. Beeler, J.A., B. Prendergast, and X. Zhuang, *Low amplitude entrainment of mice and the impact of circadian phase on behavior tests.* Physiol Behav, 2006. **87**(5): p. 870-80.
7. Chaudhury, D. and C.S. Colwell, *Circadian modulation of learning and memory in fear-conditioned mice.* Behav Brain Res, 2002. **133**(1): p. 95-108.
8. Lupi, D., et al., *The acute light-induction of sleep is mediated by OPN4-based photoreception.* Nat Neurosci, 2008. **11**(9): p. 1068-73.
9. McLennan, I.S. and J. Taylor-Jeffs, *The use of sodium lamps to brightly illuminate mouse houses during their dark phases.* Lab Anim, 2004. **38**(4): p. 384-92.
10. Peirson, S.N., et al., *Mammalian photoentrainment: results, methods, and approaches.* Methods Enzymol, 2005. **393**: p. 697-726.
11. Navara, K.J. and R.J. Nelson, *The dark side of light at night: physiological, epidemiological, and ecological consequences.* J Pineal Res, 2007. **43**(3): p. 215-24.

Assessing mouse behaviour throughout the light/dark cycle using automated in-cage analysis tools

Gareth Banks[1], Ines Heise[1], Abraham Acevedo[1], J Douglas Armstrong[2,3], Rasneer Sonia Bains[1], Heather Cater[1], Agis Chartsias[2], Stuart Peirson[4], Duncan Sneddon[1], Rowland Sillito[3], Valter Tucci[5], Sara Wells[1], Patrick Martin Nolan[1]

[1]MRC Harwell, United Kingdom;
[2]Actual Analytics Ltd., Edinburgh
[3]School of Informatics, University of Edinburgh
[4]Sleep and Circadian Neuroscience Institute, University of Oxford
[5]Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genova, Italy

An important factor in reducing variability in mouse test outcomes has been to develop assays that can be used for continuous automated home-cage assessment. Our experience has shown that this has been most evidenced in long-term assessment of wheel-running activity in mice. Historically, wheel-running in mice and other rodents has been used as a robust assay to determine, with precision, the inherent period of circadian rhythms in mice. Furthermore, this assay has been instrumental in dissecting the molecular genetic basis of mammalian circadian rhythms. In teasing out the elements of this test that have determined its robustness - automated assessment of an unforced behaviour in the home cage over long time intervals - we and others have been investigating whether similar test apparatus could be used to accurately discriminate differences in distinct behavioural parameters in mice. Firstly, using these systems, we explored behaviours in a number of mouse inbred strains to determine whether we could extract biologically meaningful differences. Secondly, we tested a number of relevant mutant lines to determine how discriminative these parameters were. Our findings show that, when compared to conventional out-of-cage phenotyping, a far deeper understanding of mouse mutant phenotype can be established by monitoring behavior in the home cage over one or more light:dark cycles.

Circadian rhythms and sleep

Several mouse mutants identified by us in circadian wheel-running screens show altered circadian period in constant dark conditions but also show alterations advances or delays in activity phase relative to the light:dark cycle (abnormal phase angle of entrainment). We have used some of these mutants for further behavioural characterisation and established that an altered phase angle of entrainment may have associated behavioural and sleep disturbances. In an effort to identify mouse mutants with disturbed sleep patterns, we have made use of an automated tracking system using infrared video monitoring of individuals over 5 days in home cages with subsequent analysis using commercially available software. We are investigating the applications of this approach further by comparing behavioural parameters in inbred strains, monitoring decline in sleep quality with ageing and using the approach as a first-pass screen for mouse mutants with sleep disturbances.

Motor function

In attempts to develop alternative tests to the confounding RotaRod test, we developed, as part of a European consortium, an automated wheel-running apparatus that assesses multiple motor parameters in mice over several weeks in the home cage. We have shown that the system offers a reliable, robust and reproducible test for assessing motor function in mice as shown by an excellent cross validation across research groups. The system measures multiple parameters over several weeks in the home cage and, with any mutant we have tested so far, can detect even early onset and/or subtle deficits in motor function. It is particularly encouraging to note that this system can be used to detect motor dysfunctions in two widely-used models of neurodegenerative disease at stages where other tests do not detect any other functional deficits. The system can be used to focus on activity during the early night phase where mice are most active and is adaptable to focusing on the most informative motor parameters for each study.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Timing and cognition

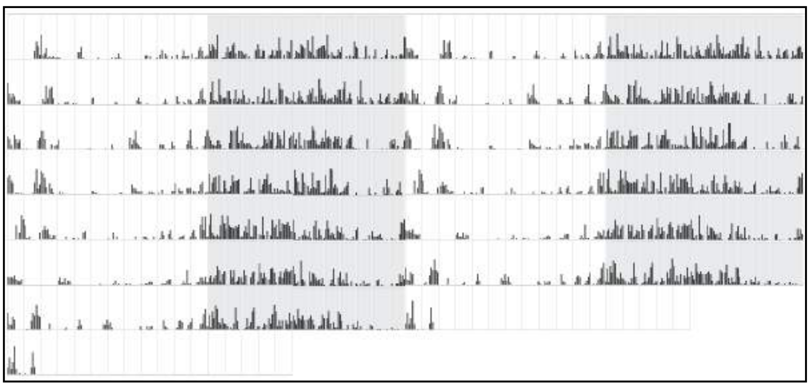
As part of a European consortium, we have been involved in the development of automated home-cage systems to study interval timing in mice over several days. The system allows us to investigate nose-poking in mice to collect food rewards over the entire daily cycle. The paradigm chosen was one where mice were trained to discriminate between the durations of two light signals (i.e., short- vs. long-latency signals) in order to obtain a food reward. The duration of the light signal determined the location of food pellet availability. By introducing probe trials throughout the task we could establish the capacity of the animal to discriminate between the two light signals. Using this apparatus, we established a robust consistent behaviour in mice that persists throughout the 24 hour cycle. Performance in this task appeared to reflect an interaction between the internal biological clock of the animal and the external environmental clock (light:dark cycle). This was confirmed using mouse mutant lines with short or long circadian clocks.

Automated assessment of individual behaviours in group-housed conditions

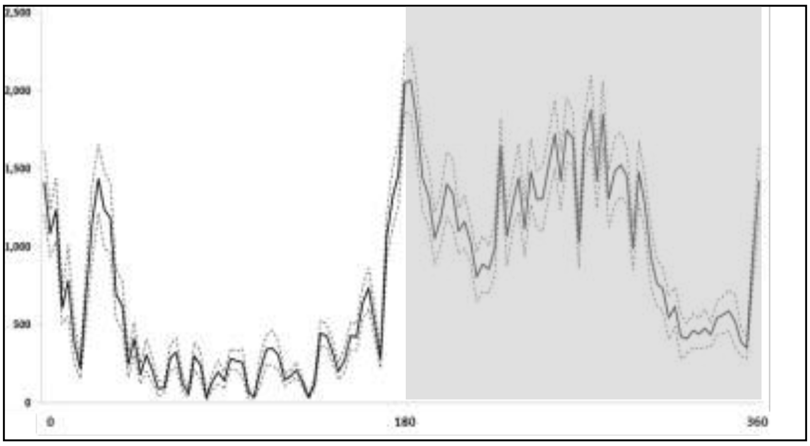
Our ability to record automated detailed behavioural parameters over time in an undisturbed cage encouraged us to explore whether true home-cage phenotyping was feasible, ie. Could we monitor the behaviours of individual animals that are reared and group-housed in conventional IVC home-cages. We were given the opportunity to test a system developed by our colleagues (see presentation by JD Armstrong at an accompanying symposium) and evaluated the system using numerous mouse strains and mutants. The system combines RFID tracking with infrared video recording for automated behavioural scoring. We focused, first of all, on activity measurements over extended periods in inbred strains. Our findings (Fig 1) show that individual strains can show distinct patterns of activity over 24-hrs in the home cage.

In one inbred strain, we found that activities are distributed throughout the 24-hr period with animals being active through the first quarter of the light phase, showing anticipatory activity prior to the onset of the dark phase and a period of inactivity during the final third of the dark phase. Furthermore, we found that many of these behaviours were strain specific. In a second strain, for example, we could not distinguish activities in the light and dark phases quite so readily. Further phenotyping allowed us to identify subtle behavioural phenotypes in mutants recorded over 24-hrs and gave us the opportunity to follow individual behaviors in cages of mixed genotype.

(a)



(b)



(c)

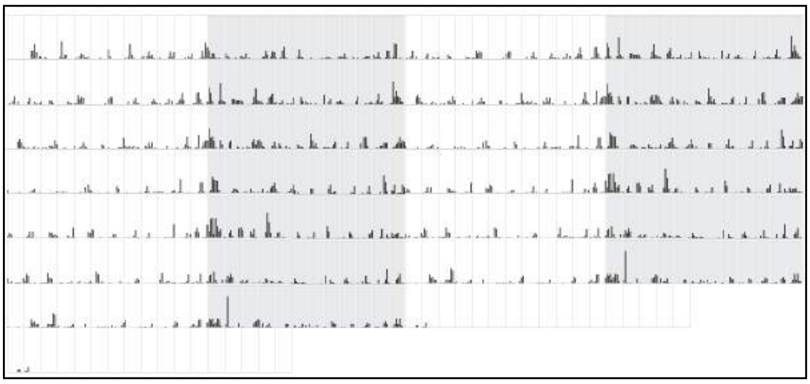


Figure 1. Home-cage activity recorded over 7 days. Home-cage activity (distance travelled) plotted in 12-minute time bins as a double-plotted raster plot (a) and a line plot (b) showing average daily activity +/- standard errors. A raster plot for a second inbred strain (c) shows differences in activity patterns among strains.

Discussion

Our findings highlight the importance of testing mouse behaviours over extended periods in undisturbed conditions (for at least one 24-hr cycle). The discriminatory power of these tests is far greater than conventional out-of-cage phenotyping. However, testing during the dark phase need not necessarily be the most informative or most discriminative. This would be dependent on the specific behaviours being studied and on the particular mouse strain or mutant line being used for the study. Careful consideration should be given to all of these factors prior to designing a behavioural study but, ultimately, longitudinal studies throughout the light:dark cycle are desirable. Progress with the automation and diversity of home-cage testing will be a critical factor in enabling such studies.

Ethical Statement

Animal studies described here were subject to the guidance issued by the Medical Research Council in *Responsibility in the Use of Animals for Medical Research* (July 1993), were dependent on an institutional Animal Welfare and Ethical Review Body evaluation and were carried out under Home Office Project Licences #30/2686, #30/2995 and #30/3206.

Statistical Analysis

Data were assessed using 1- or 2-way ANOVA depending on the protocol being used. Where applicable, Bonferroni post hoc tests were used.

Acknowledgements

This work was supported by the MRC, by the National Centre for the 3Rs (UK), by the BBSRC and by EU Programmes EUCLOCK and PHENOSCALE.

Anxiety and Fear Measures Across the Circadian Cycle: Forced versus Deliberate Responses of Mice

O. Stiedl[1,2], E. van Zwol[1], A.B. Smit[2], M, Verhage[1] and T. Hager[1,3]

[1]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research,
VU University Amsterdam, The Netherlands. oliver.stiedl@cncr.vu.nl
[2]Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research,
VU University Amsterdam, The Netherlands.
[3]Biobserve GmbH, Bonn, Germany.

Introduction

Anxiety and fear are commonly assessed in mice based on relatively short behavior tests (3-10 min) in novel environments that are performed at different times of the circadian cycle, particularly, when screening of mice in a single or few setups demands for sequential testing of mice. Commonly, tests are performed during the light phase or alternatively during the dark phase, which is favored in nocturnally active rodents. Therefore, it is important to determine whether tests at different times of the circadian cycle influence the behavioral performance, because this may contribute to the limited replicability of behavioral studies [1]. This is specifically important in tests to assess emotional states, since emotion cannot be measured directly, but instead must be inferred indirectly from physiological changes. Therefore, in rodents this is commonly based on the alteration of locomotor activity-derived measures. Performance differences as a function of the time of testing in the circadian cycle have been reported for anxiety tests with lower anxiety during the dark phase [e.g., 2] and fear conditioning experiments in mice [e.g., 3]. Passive avoidance experiments at 4-h intervals across the whole circadian cycle revealed increased avoidance during the light phase in rats [4]. Thus, circadian endocrine changes (e.g., corticosterone) can contribute to performance differences across the circadian cycle. We therefore tested the effects of the circadian phase on anxiety-like behavior and conditioned (learned) fear in two strains of mice that differ substantially in their acclaimed trait anxiety [e.g., 5], DBA/2J and C57BL/6J mice.

Material and Methods

Experiments were performed with 10-12-week-old male mice that were individually housed in several behavior assays. We used the dark/light box test to compare anxiety-like behavior, and we performed classical passive avoidance tests with forced exposure. Finally, we studied the deliberate choice of mice in an automated passive avoidance assay in a home cage environment (DualCage; see *HomeCage*^{Plus}, Biobserve GmbH, Bonn, Germany) over an extended time scale of up to 48 h to determine anxiety and learned behavior [6]. Dark/light box experiments were performed by placing mice in the dark compartment (5 lx). After a 3-min waiting period, the door was opened to the bright compartment (300 lx), and the latency to enter the bright compartment, the cumulative time in the bright compartment and the number of transitions were recorded in groups of mice tested during the light versus groups of mice tested during the dark phase. The same approach was used to assess passive avoidance learning as described before [7]. Here, training occurred by placing mice in the bright compartment (1000 lx) and opening the door to the dark compartment (10 lx). Mice were confined in the dark compartment upon their full entry and then subjected to a single shock (0.7 mA, 2 s, constant current) at 3-s delay. They were returned to their home cage 30 s after shock exposure. Fear memory was assessed 24 h later by placing the mice in the bright compartment and opening the door after 1 min for 10 min to determine the latency to enter the dark compartment, cumulative dark compartment time and number of transitions. In the DualCage a similar approach was used, except that mice were housed in the home compartment for four days before the actual experiments. Then the door to the test compartment was opened 1 h after the onset of the dark phase for the training. Upon deliberate full entry, the door was closed and with a delay of 30 s the mouse subjected to a single foot shock (0.7 mA, 2 s, scrambled current). Then the door reopened for mice to return to their home compartment. Fear memory was tested 24 h later by re-opening the door and determining the behavior of the

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

mice for 48 h without any intervention. The alleged motivations of mice to deliberately re-enter the test compartment were novelty-seeking and/or risk assessment. All animal experiments were approved by the animal research committee of the VU University Amsterdam and conducted according to Dutch regulations in compliance with the European Council Directive (86/609/EEC).

Results

We observed no difference in dark/light box performance of DBA/2J and C57BL/6J mice irrespective of the circadian cycle. Increased anxiety-like behavior of DBA/2J compared to C57BL/6J was consistently observed in our experiments. Our passive avoidance experiments also did not find performance differences between the light and the dark phase but again indicated lower avoidance, i.e., shorter transfer latencies in the retention test in DBA/2J than in C57BL/6J mice. In the DualCage the training latencies to enter the test compartment for the first time were significantly prolonged in DBA/2J mice compared to that of C57BL/6J mice. However, no difference existed in the retention latencies between the two strains with large inter-individual differences as reported by us before [7]. Only during the progression of the test compartment exploration after its first entry, which is used as an index of fear extinction [6], DBA/2J mice progressed significantly slower than C57BL/6J mice. However, what became apparent was that the deliberate decisions of mice to re-enter the test compartment occurred with high prevalence within the first half of the dark phase. The number of test compartment visits was lower in the second part of the dark phase and essentially absent in the light phase. This pattern therefore matched the activity pattern in the home cage with the activity peak in the first part of the dark phase [6]. A similar pattern was found for the time spent in the test compartment.

Discussion

We did not find differences in anxiety-like behavior in the two tested strains when comparing the performance in the dark and the light phase. Similarly, we also did not observe fear learning differences between the dark and the light phase. However, we cannot exclude such circadian performance differences if we would measure at higher temporal resolution as reported before in rats and mice [e.g., 2-4]. However, we reconfirm the increased anxiety-like behavior of DBA/2J compared to C57BL/6J mice [5]. Moreover, we extend the claim of lower fear learning of DBA/2J mice compared to C57BL/6J mice to passive avoidance learning as reported by us and others before in fear conditioning [6]. However, the DualCage experiments indicated preserved anxiety differences between DBA/2J and C57BL/6J mice but demonstrated no learning deficit, suggesting that the increased anxiety of DBA/2J mice in novel situations elicited by various stressors contributes to the observed learning deficit in this strain. Unimpaired fear learning of DBA/2J mice in the DualCage confirms earlier conclusions on the negative impact of increased emotional arousal on cognition in complex tasks [5,9].

Conclusions

In our experiments in two strains of mice we observed no difference between anxiety-like behavior and fear learning in the light versus the dark phase. However, all experiments with deliberate choice indicated that the deliberate decisions occurred largely in the first half of the dark phase. The majority of decisions correlated with the activity pattern of mice throughout the circadian cycle. Therefore, from an ethological point of view, and thus from an animal welfare aspect, it is recommendable to perform experiments that comply with the activity pattern of the animals, i.e., experiments during the first part of the dark phase of the circadian cycle in C57BL/6J mice. In addition, our experiments highlight another important issue for the interpretation of behavioral performance. Although common knowledge in animal research, it is crucial to avoid any unspecific stressors to assess intrinsically motivated emotional state and cognitive performance due to their interaction as defined by the Yerkes-Dodson law [see 10]. This can be achieved best by home cage-based phenotyping, as exploited in the DualCage, as important refinement in animal research, where the experiment comes to the animal and not vice versa.

References

1. Benjamini, Y., Lipkind, D., Horev, G., Fonio, E., Kafkafi, N., Golani, I. (2010). Ten ways to improve the quality of descriptions of whole-animal movement. Neuroscience and Biobehavioral Reviews 34, 1351-1365.
2. Bilu, C., Kronfeld-Schor, N. (2013). Effects of circadian phase and melatonin injection on anxiety-like behavior in nocturnal and diurnal rodents. *Chronobiology International* **30**, 828-836.
3. Chaudhury D., Colwell, C.S. (2002). Circadian modulation of learning and memory in fear-conditioned mice. *Behavioural Brain Research* **133**: 95-108.
4. Davies, J.A., Navaratnam, V., Redfern, P.H. (1973). A 24-hour rhythm in passive avoidance behavior in rats. *Psychopharmacologia* (Berlin) **32**, 211-214.
5. Ohl, F., Roedel, A., Binder, E., Holsboer, F. (2003). Impact of high and low anxiety on cognitive performance in a modified hole board test in C57BL/6 and DBA/2 mice. *European Journal of Neuroscience* 17, 128-136.
6. Hager, T., Jansen, R.F., Pieneman, A.W., Manivannan, S.N., Golani, I., van der Sluis, S., Smit, A.B., Verhage, M., Stiedl, O. (2014). Display of individuality in avoidance behavior and risk assessment of inbred mice. *Frontiers in Behavioral Neuroscience* **8**, 314.
7. Baarendse, P.J.J., van Grootheest, G., Jansen, R.F., Pieneman, A.W., Ögren, S.O., Verhage, M., Stiedl, O. (2008). Differential involvement of the dorsal hippocampus in passive avoidance in C57BL/6J and DBA/2J mice. *Hippocampus* **18**, 11-19.
8. Stiedl, O., Radulovic, J., Lohmann, R., Palve, M., Birkenfeld, K., Kammermeier, J., Sananbenesi, F., Spiess, J. (1999). Strain and substrain differences in context- and tone-dependent fear conditioning. *Behavioural Brain Research* **104**, 1-12.
9. Youn, J., Ellenbroek, B.A., Inti van Eck, I., Roubos, S., Verhage, M., Stiedl, O. (2012). Finding the right motivation: genotype-dependent differences in effective reinforcements of spatial learning. *Behavioural* Brain Research 226, 397-403.
10. Diamond, D. M., Campbell, A. M., Park, C. R., Halonen, J., and Zoladz, P. R. (2007). The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. Neural Plasticity 2007, 60803.

Alone at Home: Reliability of Circadian Activity Pattern Within and Between Laboratories

Gernot Riedel[1] and Berry Spruijt[2]

[1]University of Aberdeen, Institute of Medical Sciences, Aberdeen UK
[2]Utrecht University, Department of Biology, Utrecht, The Netherlands

Background

Preclinical and especially behavioural neuroscience is under increasing scrutiny due to the growing perception of irreproducibility of recorded and published data. The recent summary of Michael Jarvis and Michael Williams in TINS (2016) is the recent summary of a multitude of reports and efforts to improve this issue many of which have concentrated on the improvement and standardisation of reporting practices. These have been tackled by numerous working groups for journals and funding bodies (see Table 1 in Jarvis and Williams, 2016), and Measuring Behavior 2016 also seeks to progress towards a more detailed 'Material and Methods' section, a wider standardisation of holding conditions (open housing or individually ventilated cages or isolators), temperature neutrality of subjects during maintenance and experiments, husbandry, handling and habituation prior to experimentation, experimenter conducting the test, test protocol and equipment, etc. is required for better reproducibility of data. Towards this end, the European directive 2010/63/EU makes some progress in clearly defining new standards for housing.

In the spirit of this symposium however, a critical issue is the question of time during which experiments are conducted and whether it is appropriate for rodent experiments to be undertaken in the light phase of the cycle when rodents are typically asleep (see Hawkins and Golledge, 2016). While this typically applies to tests conducted in recording units different from the home cage, one would predict that independent whether a normal or inverted circadian rhythm is run in the holding unit, home cage observations should generate reproducible results (at least for some standard parameters such as ambulatory activity during the dark or the light phase of the cycle). Experiment 2 set out to explore this possibility in our animal facility here in Aberdeen.

At the same time, standardised home cage observation systems (for example PhenoTyper (Noldus IT, Wageningen, The Netherlands) or PhenoMaster (TSE Systems, Bad Homburg, Germany)) would be predicted to deliver similar if not identical results in different parts of the same facility and between laboratories even if scattered over different continents. That this may not be achievable using test apparatuses such as the open field and elevated plus maze or water maze (Crabbe et al., 1999) is readily explained by the non-standardisable idiosyncratic handling and test environment (background cues etc.). However, this should not apply to fully standardised home cages and the goal of Experiment 1 was to validate this proposition by keeping conditions identical between our two laboratories in Aberdeen and in Utrecht.

Material and Methods:

Male mice of the C57BL/6JOlaHsd and DBA/2OlaHsd strain from and specific-pathogen free (SBF) commercial breeder (Harlan UK) were randomly selected from a breeding colony and dispatched on the same day to the test laboratories in Aberdeen (UK; 15 mice from each strain) and in Utrecht (The Netherlands; 15 mice from each strain), where they were housed singly in an inverted 12 h day night cycle (lights on at 21.00). Water and food were provided ad libitum and all other holding conditions were equal with few exceptions (Macrolon II cages in Utrecht, shoebox in Aberdeen; enrichment was tissue (Utrecht) or wood shavings (Aberdeen). After 2 weeks of acclimatisation, animals were transferred in PhenoTyper (Noldus IT, Wageningen, The Netherlands) boxes at the same time (time difference between Netherlands and UK was adjusted) and the experiment started.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Home cage observations were performed for 1 week according to the protocol of Visser and coworkers (2006). Video tracking software Ethovision 3.1 was used to record ambulatory activity as X-Y coordinates and multiple parameters were extracted from these raw data and displayed as activity profiles over day-night cycles. Primary endpoint was the global locomotor activity in the home cage, in pre-selected zones (food zone and water zone as surrogate measures for food and water consumption) and time in shelter. Data were averaged into hourly bins or pooled together for the 12 hour dark and 12 hour light periods. Days 1-3 were considered as habituation, day 4 constituted baseline and days 5 and 6 were excluded as an additional spot-light was switched on. Statistical analysis entailed factorial analysis of variance with mouse strain and laboratory as between subject factors followed by appropriate post-hoc tests. Alpha was set to 0.05. Statistical analyses are not presented here for clarity.

Results:

Exp.1: Between-laboratory reproducibility of home cage behaviour in C57BL/6 and DBA/2 mice.

Animals were purchased from Harlan UK and delivered to the respective unit where they were acclimatised to an inverted day-night cycle for >2 weeks. Table 1 summarises the holding conditions during this period in Aberdeen and Utrecht respectively. As is obvious, there are few minor differences in holding conditions such as cage type, the enrichment used, and the fact that animals in Aberdeen were housed in the same room in which the PhenoTyper experiment took place. We projected that these minor variations should not have significant bearings on the overall experimental outcome.

The overall activity profile for Day 4 (baseline) is displayed in Fig. 1. In both strains, ambulatory activity was increased during the dark phase in Aberdeen and Utrecht. Yet, the overall activity levels during the dark phase were considerably heightened in Aberdeen relative to Utrecht (factor laboratory: $p<0.01$)

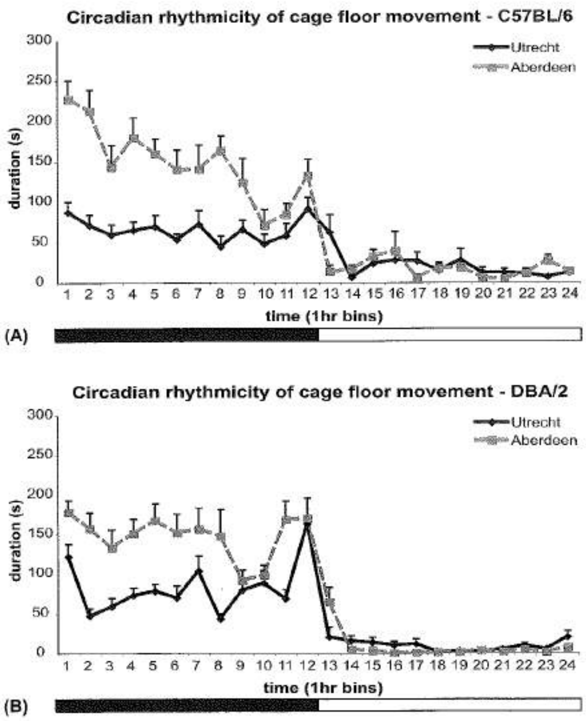


Figure 1: Circadian ambulatory activity on Day 4 during darkness (black bar) and day (open bar). Top indicates recordings for C57BL/6 mice, bottom is for DBA/2. Note that activity is heightened in the Aberdeen cohorts during the dark phase. Mean duration of movement in ho urly bins and SEM.

	Utrecht	Aberdeen
Animals		
strains	C57BL/6JOlaHsd DBA/2OlaHsd	C57BL/6JOlaHsd DBA/2OlaHsd
supplier	Harlan UK	Harlan UK
transport	by air and truck, 2 days	by truck, several hours
Housing		
cage type	Macrolon type II	"shoe box", dimension comparable with Macrolon type I elongated
bedding	Aspen chips (medium size)	Aspen chips (small-size)
enrichment	tissue	wood shavings
lighting	red: 09.00-21.00 white: 21.00 – 09.00	dark[1]: 09.00 – 21.00 white: 21.00 – 09.00
temperature	22 °C +/- 1	21 °C +/- 1
humidity	40-50 %	45-55 %
room	housed and testing in separate room	housed in same room as during home cage testing
Handling & care		
handling	once per week during cage cleaning: picked up at tail base using gloves	once per week during cage cleaning: picked up at tail base using gloves.
cage cleaning	once per week under red light	once per week under red light
food and water	twice per week	twice per week
food	SDS CRM (E)	SDS CRM (P)

Table 1: Information on animals and holding conditions at both Aberdeen and Utrecht research sites.

Exp. 2: Ambulatory activity during normal and inverted day night cycles: within subject comparison conducted in Aberdeen in DBA/2 and C57BL/6 mice.

As a follow on from the between-laboratory comparison, all mice were removed from the PhenoTypers and returned to their home cages. They were kept for over 4 weeks on a normal day night cycle and were then retested in the PhenoTypers. We reasoned that circadian activity in the home cage would be similar independent whether the day night rhythm followed the endogenous day-night-time or was inverted. Figure 2 summarises the results. There was no difference between the two mouse strains in the inverted/reverse cycle (A), but during the normal rhythm, C57BL/6 mice were less active during the night than DBA/2 (B). When considered within each strain, it became obvious that during inverted light-dark cycles, the activity was lowered in dark periods in DBA/2 mice (C), but was considerably heightened in the C57BL/6 cohort. Intriguingly, there was little difference between strains and between conditions for the light phase.

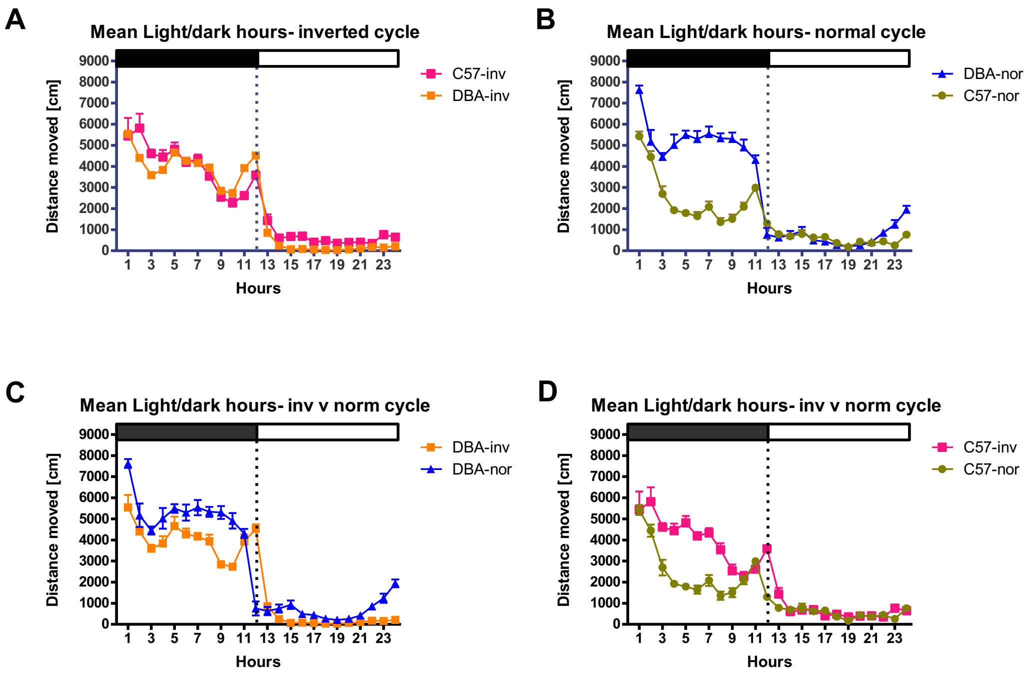


Fig. 2: Within subject comparison of circadian activity in the Phenotyper in C57BL/6 and DBA/2 mice. Two conditions were tested: (A) inverted (inv) against (B) normal (norm) day night cycle between strains; inverted against normal for DBA/2 (C) and C57BL/6 (D). Means + SEM. Black horizontal bar on top represents dark phase, open bar stands for light phase.

Conclusion:

The overarching aim was to evaluate home cage behaviour in two mouse strains in two independent laboratories. We predicted that minimising the differences between cohorts (see Table 1) would help to increase reproducibility of data. While considerable differences occur between laboratories in multiple behavioural domains (Jarvis and Williams 2016), one would hope that home cage horizontal activity in standardised home environments (here exemplified as PhenoTyper cages) is immune to such variations. 1) Animals were selected randomly from the pool of the same supplier with the same data of birth. 2) Extensive habituation to a standardised rhythm (inverted day-night cycle) was chosen prior to testing. Clearly, both laboratories reported a significant enhancement of global home cage activity during the dark phases and a low level of movements during the light phase, most likely reflecting sleep. However, considerable differences were observed, which cannot be explained in terms of equipment or settings (all identical). The most parsimonious explanation is that environmental conditions and local standards are impinging differentially on the home cage activity of cohorts. While lighting conditions remained constant, differences in environmental factors (humidity, temperature, etc.) were minimised and the housing in the same (Aberdeen) or a different (Utrecht) holding room prior to recording in PhenoTyper cages is unlikely to strongly affect the horizontal cage movements as dramatically as revealed (Fig. 1). We feel that the most likely factor that has raised the overall activity in the mice during inverted light-dark phases in Aberdeen is the fact that animals were tested in a working animal facility, in which test rooms and holding rooms are located side by side and the normal day hours (dark hours in the inverted cycle) are characterised by high activity levels of staff, experimenters and cleaning personnel. This would generate a noisy environment and arouse test animals to a higher level of activity. By comparison, the experimental unit in Utrecht comprised of a dedicated holding tract separated from the experimental rooms. Therefore, only very limited traffic and disturbance was experienced by the mice and the overall noise levels were considerably reduced relative to Aberdeen. It is unlikely in this context that an experimenter bias is responsible for such significant between-laboratory differences as animals remained unhandled throughout the testing procedure.

A corroborating experiment was conducted in Aberdeen where all mice were maintained under a normal light-dark rhythm for several weeks and then placed back into PhenoTypers for another week of home cage activity recording. While there was relatively little difference in activity between the strains in the inverted light condition, returning to a normal rhythm increased the activity level during dark in DBA/2 and reduced it in C57BL/6 mice. Thus, the already heightened ambulatory movement in the DBA/2 cohort in Aberdeen was accentuated by a shift to normal lighting conditions. Consequently, this high activity level cannot be ascribed to the global activity level within the animal unit as animals would have found it extremely quiescent during their activity phase (i.e. dark period). While this explanation still holds for the C57BL/6 strain, it can only partially explain the overall differences between strains under the two different lighting regimes.

Different levels of anxiety are not only found between strains, but also between laboratories (Crabbe et al., 1999). Towards this end, DBA/2 mice present with high levels of trait anxiety while C57BL/6 mice show specific forms of state anxiety (Riedel et al., unpublished). How these two forms may affect ambulatory activity in home cages remains to be explored in the future.

References

1. Bohlen M, Hayes ER, Bohlen B, Bailoo JD, Crabbe JC, Wahlsten D. (2014) Experimenter effects on behavioral test scores of eight inbred mouse strains under the influence of ethanol. Behav Brain Res. 272:46-54.
2. Crabbe JC, Wahlsten D, Dudek BC. (1999) Genetics of mouse behavior: interactions with laboratory environment. Science. 284(5420):1670-2.
3. Hawkins P, Golledge H. (2016) Overview of key issues associated with using nocturnal rodent during the human working day. Measuring Behavior, this issue.
4. Jarvis MF, Williams M. (2016) Irreproducibility in preclinical biomedical research: perceptions, uncertainties and knowledge gaps. Trends in Pharmacological Sciences. In Press

NMDA receptor-dependent signaling in excitatory prefrontal neurons controls fear discrimination and fear extinction

P.A. Vieira, J.W. Lovelace and E. Korzus

Department of Psychology, University of California, Riverside, CA, USA.
edkorzus@ucr.edu

Abstract. N-methyl-D-aspartate receptors (NMDARs) are implicated in synaptic plasticity and memory function including modulation of fear memory. NMDAR-dependent plasticity is widely hypothesized to be the mechanism by which memory traces are encoded and stored. Previous studies investigating fear memory focus on the amygdala and the hippocampus but the role of the medial prefrontal cortex (mPFC) in modulation of fear memory is less clear. Neurons in mPFC have been implicated in fear memory extinction. In addition, recent studies implicated prefrontal circuitry in fear memory specificity and generalization. We have recently developed a behavioral paradigm for testing fear discrimination learning. The fear discrimination task examines the ability of mice to recognize the direction (upward or downward) of frequency modulated (FM)-sweeps. This fear discrimination task requires learning of a dangerous stimulus (CS+) via classical fear conditioning (pairing with foot shock, US) and then a harmless stimulus (CS-) is associated with the nonoccurrence of the US. This assay begins with 3 days of acquisition (a single CS–US per day) followed generalization test on days 4–5. The mice are then run through discrimination training on days 7–12 in which they experience 3 sessions: first, they are tested for freezing to CS+ and CS- in context C (without any reinforcement); second, they are exposed to CS+ paired with US (or CS-); third, they are exposed to CS- (or CS+ paired with US). CS- is never reinforced through the entire procedure. During generalization and discrimination, performance was tested in context C, which is substantially different from the conditioning chamber (Context A). We tested mice with locally deleted Grin1 gene (encoding the obligatory NR1 subunit of the NMDAR) from prefrontal CamKIIα positive neurons (mPFC-NR1 KO mice) for their ability to distinguish aversive and harmless stimuli in a fear discrimination test (Fig. 1). These mice showed impaired fear discrimination following initial generalization of conditioned fear. Mice with GRIN1 deletion in excitatory prefrontal neurons exhibit normal fear responses to aversive fear-conditioned stimuli but fail to reduce their fear response to non-aversive stimuli. These data provide evidence that NMDARs in the mPFC are part of a neural mechanism supporting discriminatory fear learning. We also found that NMDAR-dependent signaling in the mPFC is crucial for fear extinction of auditory conditioned stimuli. These studies suggest that the mPFC is required for a reduction of generalized fear to harmless stimuli that is essential for improvement of fear memory accuracy. These studies have clinical implications. Overgeneralized fear is a typical symptom of anxiety disorders including generalized anxiety disorder and posttraumatic stress disorder (PTSD), which are triggered by cues in a secure environment that resemble those of the traumatic experience.

Subjects: The UC Riverside Institutional Animal Care and Use Committee approved all procedures in accordance with the NIH guidelines for the care and use of laboratory animals. We used C57BL/6J mice for all experiments. Mice were weaned at postnatal day 21, housed 4 animals to a cage with same sex littermates with ad libitum access to food and water and maintained on a 12 h light/dark cycle. Old bedding was exchanged for fresh autoclaved bedding every week.

Keywords: NMDA receptors, prefrontal cortex, fear memory, fear discrimination, PTSD, fear generalization, excitatory neurons

Figure 1. Tone fear discrimination learning is deficient in mPFC-NR1 KO mice. (A) Pavlovian tone fear conditioning was normal in mPFC-NR1 KO, showing normal acquisition of FM-sweep fear conditioning compared to control (Ctrl) mice. FM-sweep fear was tested in context C at 24 h after a three upsweeps-foot shock pairing. Shown here fear baseline is% freezing recorded before fear conditioning (day 1). (B) Both groups (mPFC-NR1 KO and Ctrl) show no difference in the freezing responses to CS+ and CS- ($p > 0.05$) during day 4 and 5 of training, indicating that initially, the mPFC-NR1 KO and control mice generalized responses and did not discriminate between CS+ and CS-. Shown here fear baseline is% freezing recorded before tone was presented during generalization test. (C) After the initial generalization of fear conditioned responses, control mice exhibited robust fear discrimination on Days 10–12. (E) mPFC-NR1 KO mice demonstrated a deficit in auditory fear discrimination when compared to controls (C). Shown here (and D) fear baseline is% freezing recorded before tone was presented during discriminatory phase (days 7–12). (E) The FM-sweep direction

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

discrimination ratios (DI) were calculated using the freezing responses to CS+ and CS_ according to the formula DI = ((\[CS+] - \[CS-])/(\[CS+] + \[CS-])). Analyses revealed differences between mPFC-NR1 KO and control mice in the performance during Days 10–12. (F) Average learning curves for learning of appropriate responses to CS+ and CS_ were calculated based on the performance of control and mPFC-NR1 KO group across the entire training (G; Days 7–12) followed by fitting the regression line and t-test analysis on the mean of those slopes (a). The analysis of patterns of responses to CS+ and CS- in control animals tested on the FM-sweep direction fear discriminatory task revealed that the improvement of auditory fear memory accuracy was due to slight incline in freezing to CS+ and rapid decline in freezing to CS- (CS+/ Ctrl: a = 2.05 ± 1.22; CS_/Ctrl: a = _5.17 ± 0.79). There was no difference in the learning (slopes) of appropriate responses to CS+ between mPFC-NR1 KO and control groups. The mPFC-NR1 KO group, which failed to improve fear memory accuracy, showed a positive slope for CS-, a marked difference from control responses to the CS-. (H) Graph showing average slopes on the same analysis. Data presented in B–F were acquired in context C during ‘‘Test’’ (D). The asterisks indicate statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001. (G) FM-sweep direction fear discrimination training protocol. (A) Schematic of the auditory fear discrimination protocol (see Methods). Following 3 days of paired FM sweep-foot shock training, a generalization test assessed the levels of freezing in response to the non-reinforced stimulus compared to the conditioned stimulus. Generalization was followed by 6 days of discrimination training (Days 7–12) in which each day consists of 3 sessions: a non-reinforced test of freezing to CS+ and CS- (‘‘Test’’), a CS+ and US pairing, and a non-reinforced CS_ session (CS+ and CS- order was counterbalanced). The conditioned stimuli (CS) for auditory fear conditioning were 20 s trains of FM-sweeps for a 400 ms duration, logarithmically modulated between 2 and 13 kHz (upsweep) or 13 and 2 kHz (downsweep) delivered at 1 Hz at 75 dB. ‘‘Test’’ indicates performance assessment. During ‘‘Test’’ 3 CS+ and 3 CS_ (30 s each) were presented after 180 s freezing baseline recording and with 180 s ITI. The assignment of the stimuli to CS+ or CS- (and order) was counterbalanced between subjects.

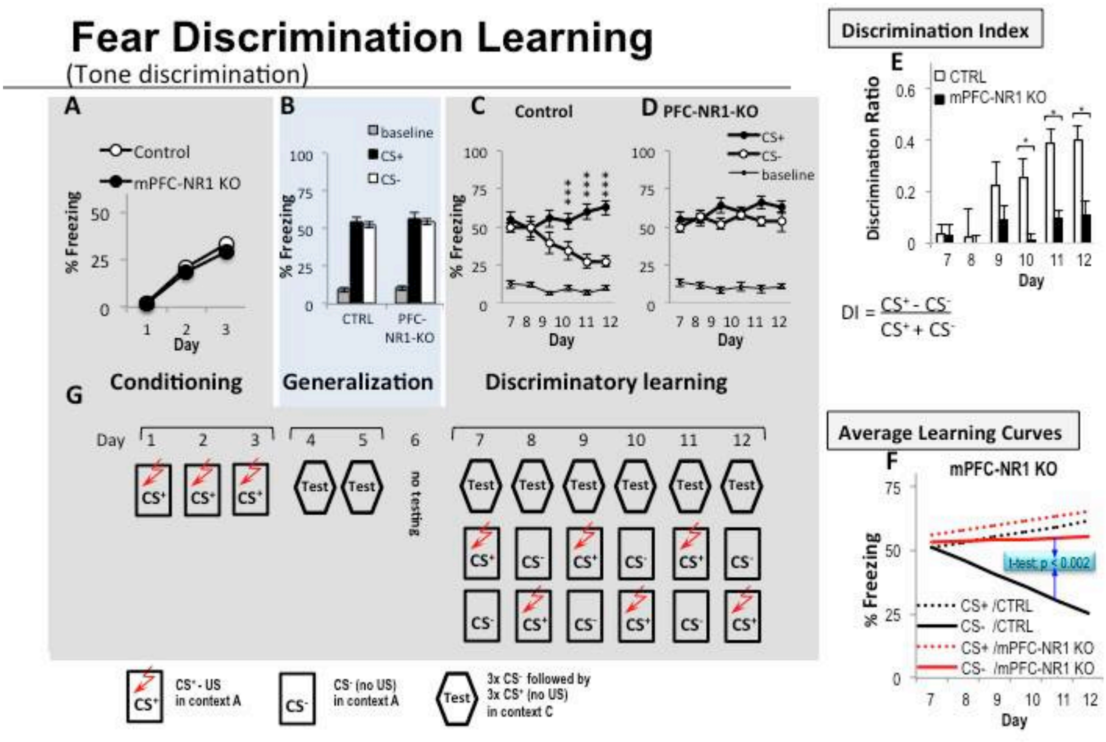


Figure 1

Amperometric Biosensors for Real-time Neurochemical Monitoring in Behaving Animals

J.P. Lowry

Department of Chemistry, Maynooth University, Co. Kildare, Ireland,John.Lowry@nuim.ie

Introduction

Molecular O_2 was one of the first substances detected electrochemically *in vivo*, both in brain [1,2] and peripheral tissues [3], using noble metal and carbon electrodes [4-7]. Brain tissue O_2 is one of the most important substrates for energy metabolism delivered by the blood and responds to a range of perturbations including electrical stimulation and neuromediator release. Nitric Oxide (NO), on the other hand, is an intracellular signaling molecule that performs a variety of roles throughout the body. It regulates vascular tone, acts as a neuronal signal in the gastrointestinal tract and central nervous system, and contributes to the pathology of several diseases including hypertension, Parkinson's disease and Alzheimer's disease. In the late 90's evidence began to appear suggesting a role for NO in modulating brain energy metabolism [8]. Around this time we had just begun to apply biosensors (e.g. glucose and blood flow) [9,10] to investigate a new model of metabolism (the astrocyte-neuron-lactate shuttle (ANLS) theory) being proposed by Magistretti and co-workers [11]. In order to obtain a more complete picture and understanding of metabolism we decided to develop new sensors to enable real-time monitoring of tissue O_2 and NO. These were based on carbon paste and polymer-modified Pt electrodes respectively. Details with respect to their development and construction are given in the Methods below, while in the Results section their characterisation and subsequent application are outlined, not only in studies of metabolism, but also in behaviour and animal models of disease (e.g. schizophrenia).

Methods

Carbon paste electrodes (CPEs) for monitoring tissue O_2 were made from 8T (200-µm bare diameter, 250-µm coated diameter) Teflon-coated silver wire (Advent Research Materials, Suffolk, UK). The Teflon insulation was slid along the wire to create an approximately 2-mm deep cavity which was packed with carbon paste using a bare silver wire as plunger. Carbon paste was prepared by thoroughly mixing 2.83 g of carbon powder (UCP-1-M, Ultra Carbon Corp., Bay City, MI, USA) and 1.0 mL of silicone oil (Aldrich, Catalogue No. 17,563-3).

Nafion®(5/2)-coated Pt disk electrodes for NO monitoring were made from Teflon®-insulated platinum/iridium (Pt/Ir 90%/10%) wire (5T, Advent Research Materials). The electrodes were approximately 4 cm in length and were prepared by carefully cutting 2 mm of Teflon® insulation from one end of the wire and soldering to this end a gold clip which provided rigidity and electrical contact. The other end of the wire acted as the active (disk) surface. This surface was then modified with Nafion® as previously described [12,13]. Experiments *in vitro* were performed in a standard three electrode glass electrochemical cell containing PBS [see e.g. 13]. For *in vivo* experiments male Sprague–Dawley or Wistar rats weighing 200–350 g were anesthetised and implanted with either O_2 or NO sensors following protocols previously described [14,15]. All procedures were reviewed and approved by the Maynooth University Biomedical and Life Sciences Research Ethics Sub-committee, and were performed under license in accordance with National (SI 543/2012) and European legislation (Directive 2010/63/EU).

Results and Discussion

In vivo characterisation of CPEs operating amperometrically at a constant potential of −650 mV (*vs.* SCE) validated *in vitro* results [16] and confirmed that they can be used reliably to monitor tissue O_2 in brain extracellular fluid, and offer several advantages compared to pulsed amperometric methods including simple experimental design, continuous real-time recording and low background noise. *In vitro* and *in vivo*

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

characterisation studies indicate a low limit of detection, high sensitivity and interference free signals at physiological concentrations, rapid response time, no effect of pH, temperature, and ion changes, no dependence upon flow (stirring), and long-term stability in vivo extending over weeks, with minimal consumption of O_2. In studies of energy metabolism where CPEs were combined with glucose and blood flow sensors we have reported an uncoupling of O_2 and glucose utilisation during neuronal activation (see Figure 1A and 1B) suggesting some level of compartmentalisation [9]. More recently, we have demonstrated that the tissue $[O_2]$ monitored using CPEs can serve as an index of changes in the magnitude of the blood oxygenation level dependent (BOLD) functional magnetic resonance imaging (fMRI) response [17]. The amperometric O_2 signal thus provides a reliable awake animal surrogate of human fMRI experimentation, and is thus an effective translational tool which can better enable the comparison of pre-clinical and clinical research. This has been substantiated by behavioural experiments where changes in reward-related CPE O_2 signals in the rat nucleus accumbens were found to be consistent with fMRI BOLD responses in man [18].

Similar *in vitro* and *in vivo* characterisation studies of the NO sensor reveal sensitive and selective detection of NO at +900 mV (*vs.* SCE). Results indicate enhanced NO sensitivity compared to bare Pt, negligible interference from potential endogenous interferents, and good operational characteristics in terms of response time and stability [12-15]. Subsequent pharmacological studies suggested that NO might play an important role in the action of the psychotomimetic NMDA-receptor antagonist phencyclidine (PCP) [19,20]. PCP was found to induce a dose-dependent increase in brain NO levels, thus corroborating previous indirect evidence of this effect of PCP. In addition, results also indicated that the PCP-induced increase in NO levels can be attenuated by the NOS-inhibitor L-NAME, in a dose that has been shown to be effective in ameliorating the schizophrenia-related behavioural effects of PCP, thus suggesting that the ability of L-NAME to block PCP-induced behaviour is directly related to an additive effect on NO levels by the two compounds. This highlights the potential of the NO pathway as a potential future target for novel treatment rationales in schizophrenia. See Figure 1C for typical *in vivo* data.

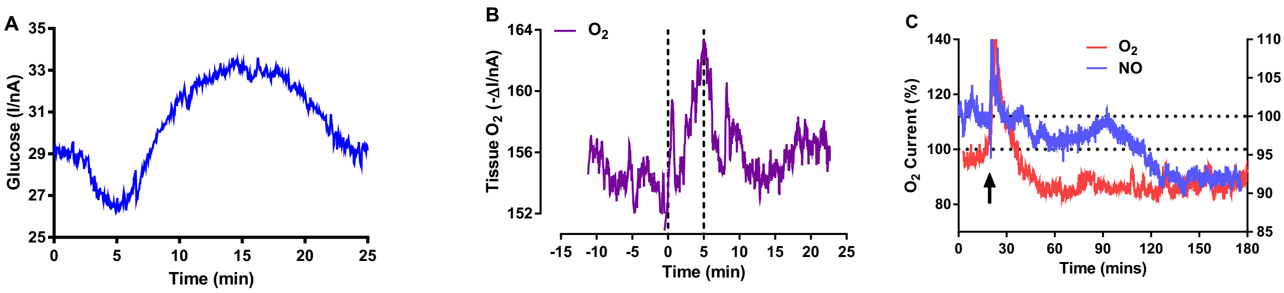


Figure 1: Typical examples of raw data obtained for the measurement of glucose (A) and O_2 (B), monitored with implanted biosensors in rat striatum in response to a 5-min tail pinch (hashed lines). (C) Simultaneously recorded rat striatal O_2 and NO data showing the response to injection (arrow) of L-NAME (30 mg/kg). Data shown as % change to facilitate comparison of μM (O_2) and nM (NO) signals.

Conclusion

Measuring real-time chemical events in the living brain is a supreme technical challenge which is now being addressed using implanted microelectrochemical sensors and *in-vivo* electrochemistry. By implanting a sensor in a specific brain region, applying a suitable potential and recording the resulting Faradaic current, changes in the concentration of a variety of substances in the ECF can be monitored with sub-second time resolution over extended periods. This is now allowing comprehensive investigations of the functions and roles of specific neurochemicals in neuronal signaling, drug actions, and behaviour.

References

1. Clark, L.C. Jr., Misrahy, G., Fox, R.P. (1958) Chronically implanted polarographic electrodes. J. Appl. Physiol., 13: 85-91.
2. Clark, L.C. Jr., Lyons, C. (1964) Studies of a glassy carbon electrode for brain polarography with observations on the effect of carbonic anhydrase inhibition. Ala. J. Med. Sci., 2: 353-359.
3. Clark, L.C. Jr., Clark, E.W. (1964) Epicardial oxygen measured with a pyrolytic graphite electrode. Ala. J. Med. Sci., 1: 142-148.
4. Lowry, J.P., Miele, M., O'Neill, R.D., Boutelle, M.G., Fillenz, M. (1998). An amperometric glucose-oxidase/poly(o-phenylenediamine) biosensor for monitoring brain extracellular glucose: in vivo characterisation in the striatum of freely-moving rats. J. Neurosci. Methods, 79: 65-74.
5. Luebbers, D.W. (1996) Oxygen electrodes and optodes and their application in vivo. Adv. Exp. Med. Biol., 388: 13-34.
6. Dittmar, A., Mangin, S., Ruban, C., Newman, W.H., Bowman, H.F., Dupuis, V., Delhomme, G., Shram, N.F, Cespuglio, R., Jaffrezic-Renault, N., Roussel, P., Barbier, D., Martelet, C. (1997) In vivo and in vitro evaluation of specially designed old and carbon fiber oxygen microelectrodes for living tissues. Sensor Actuat. B - Chem., 44: 316-320.
7. Jedlinska, B., Mellstroem, A., Hartmann, M., Joensson, K. (1998) Comparison of tissue oxygen-tension measurements by different devices. An experimental study in pigs. Scand. J. Clin. Lab. Invest., 58: 63-71.
8. Bolaños, J.P., Almeida, A. (2007) Nitric oxide in regulation of mitochondrial function, respiration, and glycolysis. In: Lajtha, A., Gibson, G.E., Dienel, G.A. (eds) Handbook of Neurochemistry and Molecular Neurobiology: Brain Energetics – Integration of Molecular and Cellular Processes. 3rd ed. Springer, New York.
9. Lowry, J.P., Fillenz, M. (1997) Evidence for uncoupling of oxygen and glucose utilisation during neuronal activation in rat striatum. J. Physiol. (London), 498: 497-501.
10. Lowry, J.P., O'Neill, R.D., Boutelle, M.G., Fillenz, M. (1998) Continuous monitoring of extracellular glucose concentratins in the striatum of freely moving rats with an implanted glucose biosensor. J. Neurochem., 70: 391-396.
11. Pellerin, L., Magistretti, P.J. (1994) Glutamate uptake into astrocytes stimulates aerobic glycolysis: a mechanism coupling neuronal activity to glucose utilization. Proc. Natl. Acad. Sci. USA, 91: 10625-10629.
12. Brown, F.O., Lowry, J.P. (2003) Microelectrochemical Sensors for In Vivo Brain Analysis: An Investigation of Procedures for Modifying Pt Electrodes Using Nafion®. Analyst, 128: 700-705.
13. Brown, F.O., Finnerty, N.J., Lowry, J.P. (2009) Nitric oxide monitoring in brain extracellular fluid: Characterisation of Nafion-modified Pt electrodes in vitro and in vivo. Analyst, 134: 2012–2020.
14. Finnerty, N.J., O'Riordan, S.L., Brown, F.O., Serra, P.A., O'Neill, R.D., Lowry, J.P. (2011) In vivo characterisation of a Nafion®-modified Pt electrode for real-time nitric oxide monitoring in brain extracellular fluid. Anal. Methods, 4: 550-557.
15. Finnerty, N.J., O'Riordan, S.L., Pålsson, E., Lowry, J.P. (2012) Brain nitric oxide: Regional characterisation of a real-time microelectrochemical sensor. J. Neurosci. Methods, 209: 13-21.
16. Bolger, F.B., McHugh, S.B., Bennett, R., Li, J., Ishiwari, K., Francois, J., Conway, M.W., Gilmour, G., Bannerman, D.M., Fillenz, M., Tricklebank, M., Lowry, J.P. (2011) Characterisation of carbon paste electrodes for real-time amperometric monitoring of brain tissue oxygen. J. Neurosci. Methods, 195: 135-142.
17. Lowry, J.P., Griffin, K., McHugh, S.B., Lowe, A.S., Tricklebank, M., Sibson, N.R. (2010) Real-time electrochemical monitoring of brain tissue oxygen: a surrogate for functional magnetic resonance imaging in rodents. Neuroimage, 52: 549-55.
18. Francois, J., Conway, M.W., Lowry, J.P., Tricklebank, M., Gilmourm G. (2012) Changes in reward-related signals in the rat nucleus accumbens measured by in vivo oxygen amperometry are consistent with fMRI BOLD responses in man. Neuroimage, 60: 2169-2181.
19. Pålsson, E., Lowry, J., Klamer, D. (2010) Information processing deficits and nitric oxide signalling in the phencyclidine model of schizophrenia. Psychopharmacology (Berl), 212: 643-51.
20. Pålsson, E., Finnerty, N., Fejgin, K., Klamer, D., Wass, C., Svensson, L., Lowry, J.P. (2009) Increased cortical nitric oxide release after phencyclidine administration. Synapse, **63**: 1083-1088.

Microdialysis and Its Significance for Measuring Behaviour

Jochen Klein

Goethe University, Germany

Microdialysis is a minimally invasive method that was developed for and is most often used in brain studies. More than 500 articles on microdialysis have been published every year since the mid-1990s, and more than 16,000 articles on microdialysis are listed in PubMed. Numerous studies in the field measured neurotransmitters in the brain, with dopamine being the most popular transmitter (about 30% of studies). In addition, other brain metabolites, especially from brain energy metabolism (glucose, lactate), have been investigated in great detail.

Microdialysis has a range of strengths and weaknesses. It is currently the preferred method to gain access to the extracellular space of the brain. Microdialysis is a sampling method (not a measuring method), therefore, the analyte(s) under study can be quantified from the dialysate using any analytical method available. Dialysates are largely protein-free which is an advantage for analysis. Other applicable methods, such as voltammetry or amperometry using sensors, have gained interest in the community, and examples of their use will be presented at the symposium. Limitations of the microdialysis technique are slow time resolutions (in the minute range), large diameters of the probe which samples from many thousands of cells and millions of synapses, and the need for corrections for recovery of analytes which is usually around 5-20%. Alternative methods with better time resolution, e.g. electrophysiology or optogenetics, will also be discussed at the symposium.

While the insertion of microdialysis probes requires anesthesia, measurements can be made in awake animals which renders the method applicable for behavioral analysis. Several examples, both form published studies [1,2] and from own work [3,4], will be presented at the symposium. For instance, several studies have investigated neurotransmitter release during behavioral activation, e.g. in the open field, or during exploration. Acetylcholine (ACh) was shown to be increased under these circumstances, and closer inspection revealed that novelty is a strong stimulus for hippocampal ACh release, and attention for cortical release [5,6]. Circadian rhythms also were accompanied by characteristic changes of transmitter release. Extensive work characterized changes of extracellular glucose, e.g. during cognitive tasks, and lactate, e.g. after arousal. A recent focus of interest was the role of neurotransmitters on feeding behavior [7]. Examples of these and other studies will be presented at the symposium.

Reference

1. Westerink, B.H.C. (1995). Brain microdialysis and its application for the study of animal behaviour. *Behav. Brain Res*. 70, 103-124.
2. Fillenz, M. (2005) In vivo neurochemical monitoring and the study of behaviour. *Neurosci. Biobehav. Rev*. 29, 949-962.
3. Hartmann, J., Kiewert C., Klein, J. (2010) Acetylcholine release and energy metabolites in amyloid-bearing APP_{SWE} x PSEN1dE9 mice. *J. Pharmacol. Exp. Ther.* 332, 364-370.
4. Kopf, S.R., Buchholzer, M.L., Hilgert, M., Löffelholz, K., Klein, J. (2001) Glucose plus choline improve passive avoidance behaviour and increase hippocampal acetylcholine release in mice. *Neuroscience* 103, 365-371.
5. Pepeu G., Giovannini M.G. (2004) Changes in acetylcholine extracellular levels during cognitive processes. *Learn. Mem*. 11, 21-27.
6. Gold, P.E. (2003) Acetylcholine modulation of neural systems involved in learning and memory. *Neurobiol. Learn. Mem*. 80, 194-210.
7. Hernandez, L., Paredes, D., Rada, P. (2011) Feeding behavior as seen through the prism of brain microdiaysis. *Physiol. Behav*. 104, 47-56.

Dietary Behavior: Progress to Date and Future Directions for Dietary Assessment

A.L. Yaroch and T.M. Smith

Gretchen Swanson Center for Nutrition, Omaha, USA.
ayaroch@centerfornutrition.org

Introduction

The increasing rates of obesity in the United States (US), as well as worldwide continue to be problematic and affect multiple populations, such as children, socioeconomically disadvantaged populations, and ethnic groups [1–3]. All overweight and obese Americans are at risk for obesity-related chronic diseases, such as diabetes mellitus, hypertension, and certain types of cancers [4]. Poor diet and physical inactivity are key determinants of obesity. Assessing and characterizing dietary intake is an essential first step to promoting a more healthful diet and ultimately reducing obesity.[5]

Health behaviors are complex and typically difficult to assess but dietary behavior and assessment techniques have not evolved sufficiently for researchers and practitioners. For instance, accelerometers exist as a strong objective technique to assess physical activity behavior but there is no comparable objective assessment for diet. Although numerous dietary assessment methods are available, it is often challenging for researchers, practitioners, and evaluators to select the best method due to inherent limitations associated with all measures. More comprehensive methods include 24-hour dietary recalls and food records/diaries, which tend to be burdensome and expensive but are able to reasonably assess daily energy and nutrient intakes. Other briefer methods such as dietary screeners are short, easy to administer, and inexpensive but are only able to assess "patterns" as a proxy of dietary behaviors.

This session will describe the various dietary assessment methods available and will highlight associated strengths and limitations, as well advances in technology. Specifically, innovations such as the Automated Self-Administered 24-hour recall (ASA-24) (versions for children and adults), as well as briefer instruments such as the 2009-10 NHANES Dietary Screener will be discussed. Critical online databases will also be explored, including the National Collaborative on Childhood Obesity Research (NCCOR) Measures Registry and Catalogue of Surveillance Systems. Recommendations of appropriate dietary assessment tools will be provided for diverse audiences, based on various behavioral intervention study designs and more practitioner-based projects.

Objective Dietary Assessment

Dietary assessment generally falls within two methodologies: objective and subjective. Objective dietary assessments, strictly speaking, are highly resource intensive and burdensome. However, they also tend to generate the most robust data. Subjective dietary assessments range in their intensity and burden, as well as the rigor of outcome data.

The criterion standard measures of diet includes two intensive laboratory methods: doubly labeled water and indirect calorimetry [6, 7]. The doubly labeled water technique tracks carbon dioxide and water produced during physical activity, in comparison to calories consumed. This method is time intensive as it generally takes 4 to 21 days, capturing habitual energy consumption and expenditure patterns. This technique has been considered a criterion standard for over 30 years; however, it is also typically costly and time prohibitive [8]. Indirect calorimetry is a technique that estimates energy expenditure from measures of carbon dioxide production and oxygen consumption during rest and steady-state exercise. Indirect calorimetry is carried out on an individual basis, which makes it fairly time-intensive leading to high participant burden. These methods have often been used to validate less intensive methods [8]. However, they are not recommended in community settings.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Subjective Dietary Intake Assessment Methods

Subjective dietary intake measures are more feasible for implementation than the criterion standard measures and can be tailored to clinician, population researcher, and practitioner needs. Subjective dietary assessments rely on self-report data to varying extents and often employ other measures to improve accuracy. While many subjective dietary intake assessments are highly developed and tested, longer methods tend to also be resource intensive with results that may not truly represent actual intake [8]. Traditional approaches include 24-hour dietary recalls, diet records, diet histories, food frequency questionnaires (FFQs), and dietary screeners [8]. Novel methods are available and/or are in development, which are potentially less resource intensive, but may not be as well developed or tested [8]. When choosing a dietary assessment method, it should first be determined what type of dietary information is optimal and/or feasible to collect (e.g., caloric and nutrient intake, specific food groups, recent or long-term intake, etc.) and then choose the appropriate tool, weighing in the pros and cons [8]. Table 1 provides an overview of subjective dietary assessment methods.

Method	Outcome	Pros	Cons
Screeners	Dietary patterns	- Less subject and researcher burden - Can show changes (food categories, compare groups)	- Intake is not quantifiably precise - Relies on subject recall
Food frequency questionnaire	Habitual intake	- Less subject and researcher burden - Can show changes (food categories, compare groups)	- Intake is not quantifiably precise - Relies on subject recall
Diet Record	Quantifiable	- Can identify energy and micro- and macro-nutrients - Does not require recall	- High subject and researcher burden - Can alter eating behavior - Need highly literate population
Multiple 24-hour recall	Habitual intake, quantifiable	- Does not require literate population	- Can have high researcher burden - Relies on subject recall

Table 1. Subjective dietary assessment methodologies.

Subjective Dietary Assessment Methods: Gold Standards

There are two types of subjective dietary assessment that are generally regarded as the gold standard: diet records and 24-hour dietary recalls. The diet record involves the subject recording all foods and beverages consumed during a specified time period, usually for 3, 5, or 7 days [9]. Because of this commitment, this method is considered time intensive and burdensome for both the subject and the clinician, researcher, or practitioner [9]. The subject is also expected to be highly motivated and literate [8]. Traditionally, the diet record is tracked using pen-and-paper, but a novel adaptation is the digital image-based food record (DIFR), which photographs are taken of food and beverage items before and after eating. The DIFR transfers burden from the subject to the researchers, as photos are then analyzed using visual aids to help judge portion size and estimate the amount of food and beverage that was consumed, and has been found to produce reliable results similar to a traditional diet record [10]. Younger generations may prefer the DIFR because it utilizes interactive technology [8]. Another novel method is the Healthy Eating Self-Monitoring Tool (HEST), which is a computer-mediated diet record designed to specifically assess fruit and vegetable intake [11, 12]. The HEST has been validated in a sample of African American adolescents and was determined to be a viable alternative to a traditional diet record [11, 12].

The other gold standard for subjective dietary assessment is the 24-hour dietary recall, which is a structured interview where a trained professional or automated program guides the respondent to list all foods and beverages consumed during the previous 24 hours [8]. This method requires only a moderate amount of time to complete and has decreased participant burden when compared to diet history or diet record [9], as it only requires the respondent to recall foods and drinks consumed in the past day. It is also generally believed to cause less response bias due to social desirability. In order to be representative of normal dietary patterns, it is recommended that at least 3 recalls be collected to obtain estimates of usual individual-level intake [13]. The Automated Self-Administered 24-Hour Recall (ASA24) is a web-based 24-hour dietary recall that asks respondents about meals consumed in the past 24 hours, contains probes for foods and beverages consumed between meals, as well as obtains information on details about each food and beverage (e.g., serving size and preparation method), forgotten items, and usual intakes. ASA24 data can be collected in English and Spanish and is available for use free of charge for researchers, clinicians, practitioners and others [14].

Subjective Dietary Assessment Methods: Other Methods

The diet history method involves a trained interviewer asking respondents to report usual food intake over an extended period of time. The interviewer provides cognitive support to probe for food preparation methods and intake patterns. This method can be highly burdensome and resource intensive [9] since the respondent must cognitively recall foods and drinks eaten over several days in the past.

Food frequency questionnaires (FFQ) assess frequency of consumption of particular foods and beverages for a specified time period. They can be interviewer or self-administered by various modes, such as scannable response form, paper and pencil, internet, etc. [9]. The FFQ provides minimal information when compared to the gold standard methods, but it has the added advantage of relatively low participant burden [9] and can focus on specific food groups or patterns of interest. The FFQW82 is a novel, self-administered 82-item FFQ that assesses intake frequency and portion size over the period of one month from a list of pictured foods and can be completed in 30 minutes [15]. The FFQW82 was adapted for an adolescent population and was determined to be reliable and valid for estimating energy and nutrient intake for breakfast, lunch, and dinner, as well as the whole day, compared with a diet record, when evaluated among Japanese females ages 12 to 13 years [15]. For clinicians, the Patient-centered Assessment and Counseling for Exercise plus Nutrition (*PACE+*) is a novel, computer-based FFQ that also assesses physical activity and can be administered in a waiting room[16]. In 2006, the *PACE+* program was tested in health care settings and found to be acceptable by participants [16, 17].

Dietary screeners are brief surveys that typically obtain information about patterns of behaviors or focus on a particular food group (such as fruits and vegetables, fast foods, various types of beverages, etc.). They do not estimate calories and nutrient values but rather obtain gross level estimates of intake for a particular food/beverage or food group. Screeners can be useful when patterns of dietary behaviors are of interest, but when limited time and/or funds are available. The 26-item Dietary Screener Questionnaire (DSQ) from the 2090-2010 National Health and Nutrition Examination Survey (NHANES) asks about frequency of intake of fruits and vegetables, dairy/calcium, added sugars, whole grains/fiber, red meat, and processed meat in the past month [18–21]. The DSQ provides a short, relatively simple way to rank adults and older children with regard to consumption of certain foods and nutrients. It provides accurate estimates of food groups and dietary patterns, which will allow for examination of interrelationships between diet and other variables in this study [19–22]. Several items from the DSQ have been tested for validity against multiple 24-hour recalls and all 26 items have been cognitively tested.[21] The DSQ, which is publicly available online (http://appliedresearch.cancer.gov/nhanes/dietscreen/questionnaires.html) and is accompanied by a SAS program that can be utilized to analyze the data. The scoring algorithms convert screener responses to estimates of dietary intake for fruits and vegetables (cup equivalents), dairy (cup equivalents), added sugars (tsp), whole grains (ounce equivalents), fiber (g), and calcium (mg). Red meat and processed meat questions will be used as qualitative indicators of intake frequency [21]. Results can be compared to national estimates [21]. The 2009-2010 NHANES dietary screener was adapted for use in the Family Life, Activity, Sun, Health, and Eating (FLASHE) Study, a new National Cancer Institute (NCI) survey on the psychosocial, generational (parent-adolescent (12-17 years)), and environmental correlates of cancer preventive behaviors. The screener involves common items (e.g., green salad, non-fried vegetables, cooked beans, fruit, cooked whole grains, whole grain bread, non-sugary cereal, fried chicken, sugary cereal, candy and chocolate, fried potatoes, other non-fried potatoes) as well as novel items to capture dietary habits of adolescents and their families (e.g., chips, processed meat, burgers, tacos, heat and serve, etc.).

Subjective Dietary Assessment Methods Database

Due to the ability of dietary assessment methodologies to be tailored to fit a niche, there are a multitude of options available. Therefore, the National Collaborative on Childhood Obesity Research (NCCOR) Measures Registry and Catalogue of Surveillance Systems are critical databases for researchers. The Measures Registry provides a searchable database of diet measures relevant to childhood obesity research, which includes assessments appropriate for adults, as well. The purpose of the Measures Registry is to standardize the use of common measures and research methods across childhood obesity research at the individual, community, and population levels. The Catalogue of Surveillance Systems provides researchers an overview of a diverse array of publically available surveillance systems. Utilization of shared measures is key to advancing the field, though it is still relatively novel. The National Cancer Institute's Dietary Assessment Primer is another critical resource that acknowledges where measurement error exists, where and what kind of burden is expected, and guidance for choosing a dietary assessment methodology for specific research questions [23].

Recommendations Moving Ahead in Dietary Assessment

As one peruses the vast collection of dietary assessment methodologies currently available, it becomes clear that participant and researcher burden need to be taken into consideration for ultimate feasibility. However, as burden is reduced, the quality of the data often becomes compromised. To date, this issue has been explored, but not adequately addressed. Critics maintain that human memory is inaccurate when it comes to diet, and that when data are not independently observed, they are inadmissible in scientific research [24]. Proponents of subjective dietary assessment cite that while limitations of these methods are openly shared in research and that predictive validity of dietary recall has been repeatedly documented [25–27], accuracy can be only maintained by using validated protocols well-trained research personnel [25]. However, reducing subjectivity on the reporting side continues to be an area where research has failed to keep up with the times. Researchers, clinicians, and practitioners alike need to find innovative ways to to try and accurately capture dietary behaviors, while minimizing bias and burden. One underdeveloped avenue is the use of momentary assessment of dietary intake, such as through the use of real-time photography using a subject's smart phone. Other areas of future research exist in the expansion of databases and development of savvy data analysis methods, along with an update of the techniques themselves [23].

References

1. Ogden, C. L., Lamb, M. M., Carroll, M. D., & Flegal, K. M. (2010). Obesity and socioeconomic status in adults: United States 1988–1994 and 2005–2008. NCHS data brief no 50. National Center for Health Statistics: Hyattsville, MD.
2. Sobal, J., & Stunkard, A. J. (1989). Socioeconomic status and obesity: A review of the literature. Psychological Bulletin, 105(2), 260–275. doi:10.1037/0033-2909.105.2.260
3. Ogden CL, Carroll MD, Kit BK, & Flegal KM. (2014). Pevalence of childhood and adult obesity in the united states, 2011-2012. JAMA, 311(8), 806–814. doi:10.1001/jama.2014.732
4. Flegal, K. M., Carroll, M. D., Kit, B. K., & Ogden, C. L. (2012). Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. JAMA, 307(5), 491–497.
5. Patrick, K., Norman, G. J., Calfas, K. J., Sallis, J. F., Zabinski, M. F., Rupp, J., & Cella, J. (2004). Diet, physical activity, and sedentary behaviors as risk factors for overweight in adolescence. Arch Pediatr Adolesc Med, 158(4), 385–390. doi:10.1001/archpedi.158.4.385
6. Schutz, Y., Weinsier, R. L., & Hunter, G. R. (2001). Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures. Obesity Research, 9(6), 368–379. doi:10.1038/oby.2001.48
7. Seale, J. L., Conway, J. M., & Canary, J. J. (1993). Seven-day validation of doubly labeled water method using indirect room calorimetry. Journal of applied physiology (Bethesda, Md.: 1985), 74(1), 402–409.
8. Yaroch, A. L., Byker, C., Pinard, C. A., & Smith, T. M. (2012). Advances in methodologies for assessing dietary intake and physical activity among adolescents. Adolescent medicine: state of the art reviews, 23(3), 610–630.

9. Thompson, F. E., & Byers, T. (1994). Dietary assessment resource manual, 124(11 Suppl), 2245S–2317S.
10. Matthiessen, T. B., Steinberg, F. M., & Kaiser, L. L. (2011). Convergent validity of a digital image-based food record to assess food group intake in youth. Journal of the American Dietetic Association, 111(5), 756–761. doi:10.1016/j.jada.2011.02.004
11. Di Noia, J., Contento, I. R., & Schinke, S. P. (2007). Criterion validity of the Healthy Eating Self-monitoring Tool (HEST) for black adolescents. Journal of the American Dietetic Association, 107(2), 321–324. doi:10.1016/j.jada.2006.11.015
12. Di Noia, J., & Contento, I. R. (2009). Criterion validity and user acceptability of a CD-ROM-mediated food record for measuring fruit and vegetable consumption among black adolescents. Public health nutrition, 12(1), 3–11. doi:10.1017/S1368980008001754
13. Johnson, R. K. (2002). Dietary intake - how do we measure what people are really eating? Obesity, 10, 63S–68S. doi:10.1038/oby.2002.192
14. Subar, A. F., Kirkpatrick, S. I., Mittl, B., Zimmerman, T. P., Thompson, F. E., Bingley, C., … Potischman, N. (2012). The Automated Self-Administered 24-Hour Dietary Recall (ASA24): A Resource for Researchers, Clinicians, and Educators from the National Cancer Institute. Journal of the Academy of Nutrition and Dietetics, 112(8), 1134–1137. doi:10.1016/j.jand.2012.04.016
15. Watanabe, M., Yamaoka, K., Yokotsuka, M., Adachi, M., & Tango, T. (2011). Validity and reproducibility of the FFQ (FFQW82) for dietary assessment in female adolescents. Public health nutrition, 14(2), 297–305. doi:10.1017/S1368980010001618
16. Patrick, K., Calfas, K. J., Norman, G. J., Zabinski, M. F., Sallis, J. F., Aaseth, J., … Cella, J. (2006). Randomized controlled trial of a primary care and home-based intervention for physical activity and nutrition behaviors: PACE+ for adolescents. Archives of Pediatrics & Adolescent Medicine, 160(2), 128–136. doi:10.1001/archpedi.160.2.128
17. Patrick, K., Sallis, J. F., Prochaska, J. J., Lydston, D. D., Calfas, K. J., Zabinski, M. F., … Brown, D. R. (2001). A multicomponent program for nutrition and physical activity change in primary care: PACE+ for adolescents. Archives of Pediatrics & Adolescent Medicine, 155(8), 940–946.
18. Thompson, F. E., & Byers, T. (1994). Dietary assessment resource manual. Journal of Nutrition, (124), 2245S–2317S.
19. Yaroch, A. L., Byker, C., Pinard, C. A., & Smith, T. M. (2012). Advances in Methodologies for Assessing Dietary Intake and Physical Activity among Adolescents. Adolesc Med, (23), 610–630.
20. National Cancer Institute. (n.d.). Screeners at a Glance. Dietary Assessment Primer. Retrieved March 13, 2015, from http://dietassessmentprimer.cancer.gov/profiles/screeners/
21. National Cancer Institute, & Applied Research Cancer Control and Population Sciences. (2014, April 11). Dietary Screener Questionnaire in the NHANES 2009-10. Retrieved from http://appliedresearch.cancer.gov/nhanes/dietscreen/
22. Coulston, A. M., Boushey, C., & Ferruzzi, M. (2013). Nutrition in the Prevention and Treatment of Disease. Academic Press.
23. Thompson, F. E., Kirkpatrick, S. I., Subar, A. F., Reedy, J., Schap, T. E., Wilson, M. M., & Krebs-Smith, S. M. (2015). The National Cancer Institute's Dietary Assessment Primer: A Resource for Diet Research. Journal of the Academy of Nutrition and Dietetics, 115(12), 1986–1995. doi:10.1016/j.jand.2015.08.016
24. Archer, E., Pavela, G., & Lavie, C. J. (2015). The Inadmissibility of What We Eat in America and NHANES Dietary Data in Nutrition and Obesity Research and the Scientific Formulation of National Dietary Guidelines. Mayo Clinic Proceedings, 90(7), 911–926. doi:10.1016/j.mayocp.2015.04.009
25. Davy, B. M., & Estabrooks, P. A. (n.d.). The Validity of Self-reported Dietary Intake Data: Focus on the "What We Eat In America" Component of the National Health and Nutrition Examination Survey Research Initiative. Mayo Clinic Proceedings, 90(7), 845–847. doi:10.1016/j.mayocp.2015.05.009
26. Hu, F. B., Stampfer, M. J., Manson, J. E., Rimm, E., Colditz, G. A., Rosner, B. A., … Willett, W. C. (1997). Dietary fat intake and the risk of coronary heart disease in women. The New England Journal of Medicine, 337(21), 1491–1499. doi:10.1056/NEJM199711203372102
27. Satija, A., Yu, E., Willett, W. C., & Hu, F. B. (2015). Understanding nutritional epidemiology and its role in policy. Advances in Nutrition (Bethesda, Md.), 6(1), 5–18. doi:10.3945/an.114.007492

How to communicate Italian wine in China. Guidelines from a multi-method study by means of interview and neuromarketing tools

Andrea Cicero, Giulia Songa, Maurizio Mauri, Vincenzo Russo

Iulm University, Italy

The present study had two main aims. The first one was to investigate the perception that Chinese culture has about Italian wine. The second one, a neuroscientific study, was aimed to investigate the effect of wine communication on Chinese people, with a particular focus on the role of colours.

The study starts from the assumption that in purchasing behaviour, the emotional and indentificatory dimensions have a predominant role compared to the functional value of the product [1;2;3;4;5], which has a symbolic function related to the cultural meaning [6]. Preferences and consumption practices are partly manifestations of cultural categories [7;8;9] and this phenomenon is particularly relevant in the field of food consumption, because it is characterized by peculiar symbolic and psychological value [4;10;11] and the choice of a food product is driven especially by its experiential value, symbolic and aesthetic [1; 5; 2; 12,13].

Basing on these evidences, has been defined the study, divided into two phases.

The first phase explored the knowledge, the consumer behaviour and the perception of Italian wine on a sample of 780 Chinese (equally distributed in Shanghai, Pechino, Hong Kong, Chongqing, Tientsin, Canton), via CAWI interviews. A number of 38 questions (likert scale, 7 points) investigated different aspect about Chinese consumption of wine and perception of Italian wine.

The second phase was conducted in a laboratory of neuroscientific research where the promotional communication materials was presented to a sample of 13 Chinese subjects (6 M), detecting the cognitive and affective responses, through the recording of psychophysiological signals, the visual behaviour and explicit assessment. On the basis of the differences between individuals from East and West cultures at the perceptual level [14] and at allocation of attentional resources [15], the study focused on specific characteristics of communication advertising, such as colour and background, in order to identify guidelines for the effective communication of Italian wine on the Chinese market.

To explore this issue, we tested two version of the same printed advertising, making a changing in the colour of landscape. The landscape of the original version was yellow; the colour of the modified version was light blue (see Figure 1).

With the aim to achieve significant and objective data to draw reliable conclusions, the two versions of advertising were compared by making extensive use of recent neuroscientific methodologies, such as electromyography (EMG), skin conductance (SC) and eye tracker technology (ET).

EMG measures the level of muscular contraction of the corrugator muscles of the face, used to infer the valence attributed to the stimulus to which the subject is exposed, as was demonstrated a high correlation between this muscle activity and emotional states [16;17;18] in different fields of marketing included print advertisements [19].

SC, measured the autonomic changes in the electrical properties of the skin attributable to a stronger activation of sweat glands controlled by the sympathetic autonomic nervous system through the hypothalamus. Skin conductance have been used in a wide array of research include print advertising, serving as indicators of such processes as attention, habituation, arousal, and cognitive effort in many different subdomains of psychology and marketing related disciplines [20;21]. ET is a technology able to examine the visual behaviour of a person. Since

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

eye movements are closely related to cognitive processes, they are valid indicators to measure the attention and all the process that can result from attention, such as the appreciation or the selection processes [22; 23].

ET has been already used in different fields of marketing included adverting on newspaper. Already in the early 1900s, Nixon [24] first applied eye-movement research to determine the attention capture value of magazine and newspaper advertisements. In recent years many researches have studied visual attention to print advertising [25; 26;27].

Figure 1. Two version of the advertising tested. On the left the original version, on the right the modified version.

The combined results from these two studies have shown a very particular perception of Italian wine. For example, foreign people consider Italian wines unique due to the history and traditions transpiring from them. Taste and scent are other aspects that makes Italian wines unique.

Moreover, the study indicate the specific influence of certain characteristics of promotional stimuli on the appreciation and perception of Italian wine, noting the cultural connotation of the consumer and providing some guidelines for the communication of this product in the Chinese market. In particular, the colour used in the communication has demonstrated to be a key aspect able to generate different reaction. Yellow colour is more appreciated by Chinese people than blue one.

Reference

1. Fabris G. (2003). Il nuovo consumatore verso il postmoderno, Franco Angeli, Milano.
2. Olivero N., Russo V. (2009). Manuale di psicologia dei consumi. Individuo, società, comunicazione, McGraw-Hill, Milano.
3. Olivero N., Russo V. (2013). Psicologia dei consumi, McGraw-Hill, Milano.
4. Russo V., Marelli S., An gelini A. (2011). Consumo critico: alimentazione e comunicazione. Valori e comportamenti per un consumo sostenibile, Franco Angeli, Milano.
5. Siri G. (2001). La psiche del consumo, Franco Angeli, Milano.
6. Douglas M., Isherwood B. (1984). Il mondo delle cose, Il Mulino, Bologna.
7. Bourdieu, P. (1984). Distinction: A social critique of the judgement of taste. Cambridge, UK: Harvard

University Press.
8. Lamont, M., Molnar, V. (2001). How blacks use consumption to shape their collective identity: Evidence from marketing specialists. Journal of Consumer Culture, 1, 31–45.
9. Miller, D. (1998). A theory of shopping. Oxford: Polity.
10. Russo V (2016). Comunicare il vino. Tecniche di neuromarketing applicate. Marketing, comunicazione e vendite. Guerini Next.
11. Russo V. (2015). Neuromarketing, comunicazione e comportamenti di consumo. Principi, strumenti e applicazioni nel food and wine. Milano: Franco Angeli.
12. Ciceri A., Stasi A, Nardone G., Songa G., Mauri M., Russo V. (2015). Effect of information on food evaluation and willingness to buy. A study from a neuromarketing perspective. Neuromarketing Theory & Practice, 14, nov. 2015.
13. Ciceri A., Songa G., Chiappa M., Mauri M., Bellotto M. (2015). Wine label and consumer evaluations. Findings from an eye tracking study, in Proceedings of 14th European Congress of Psychology, Milano.
14. Masuda, T. Nisbett, R. E. (2001). Attending holistically vs. analytically: Comparing the context Sensitivity of Japanese and Americans. Journal of Personality and Social Psychology, 81, 922-934.
15. Chua, H.F., Boland, J. E. Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. Proceedings of the National Academy of Sciences, 102, 12629–12633.
16. Cacioppo, J. T., Petty, R. E., Losch, M. E., Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. Journal of Personality and Social Psychology, 50, 260–268.
17. Cacioppo, J.T., Bush, L. K., Tassinary, L. G. (1992). Microexpressive facial actions as a function of affective stimuli: Replication and extension. Personality and Social Psychology Bulletin, 18, 515–526.
18. Zurawicki L. (2010). Neuromarketing Exploring the Brain of the Consumer, Springer, NY.
19. Ravaja, N. (2004). Contributions of Psychophysiology to Media Research: Review and Recommendations. Media Psychology, 6, 193-235.
20. Bolls P., Lang A, Potter R. (2001). The effects of message valence and listener arousal on attention, memory and facial muscular responses to radio advertisements. Communication Research, 28, 627–51.
21. Russo J.E. (1978). Eye fixations can save the world: A critical evaluation and a comparison between eye fixations and other information processing methodologies, Advances in Consumer Research, Hunt, ed., 561–570, Association for Consumer Research, Ann Arbor, MI.
22. Rizzolatti G., Riggio L., Sheliga B.M. (1994). Space and selective attention, Attention and Performance, Umiltà C., Moscovitch M., eds., 231–265, MIT Press, Cambridge, MA.
23. Nixon, H. K. (1924), Attention and interest in advertising. Archives of Psychology, 72, 5–67.
24. Pieters, R., Rosbergen, E., Wedel, M. (1999). Visual attention to repeated print advertising: A test of scanpath theory, Journal of Marketing Research.
25. Pieters R., Wedel M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects, Journal of Marketing, 68, 36-50.
26. Wedel M., Pieters R. (2000). Eye fixations on advertisements and memory for brands: A model and findings, Marketing Science, 19, 297-312.

The Impact of Communication About Food on Preference and Purchase Intention: a Neuromarketing Experiment

Maurizio Mauri[1], Andrea Ciceri[1], Giulia Songa[1], Antonio Stasi[2], Gianluca Nardone[2] and Vincenzo Russo[1]

[1]IULM University of Milan, Italy
[2]Università degli Studi di Foggia

Introduction

In this work it is presented an experimental study carried out at the "Behavior & Brain Lab" of the IULM University of Milan. The experiment was realized applying both traditional and neuroscientific methods to evaluate the role of emotional responses related to food vision and food tasting. In particular, the food (an Italian breakfast snack) was presented into two ways: one way consisted of food conserved by means of traditional techniques, the other one was represented by food conserved with a new edible gel which avoids the presence of artificial preservatives. In this work we present some preliminary data about the impact of communication on the food consumption experience. Results enable to claim a significant impact of information on the sensorial experience about food, both in term of conscious and unconscious reaction during food tasting.

Scientific rationale

Nowadays the spreading out of neuroscientific knowledge and techniques in different field such as consumer psychology, communication and marketing leads to the rise of new disciplines such as Neuromarketing and Neuroeconomics. The application of neuroscientific methods, generally based on brain imaging and biological indicator monitoring, is leading to a more objective, observable and repeatable research study related to the attempt to evaluate the emotional contribute in cognitive processes of evaluation and decision making. However, there are only few studies (de Wijk et al., 2012) trying to investigate these topics in food research. Nevertheless, the limited scientific experiments showed promising results that indicate the possibility to widen the research work on this topic.

Experimental procedure

The sample of subjects who participated to the experiment were sixty students in total. they were divided according to an equal distribution in terms of age, gender and socio-economic status. All participants tasted the same Italian snack (called "saccottino") three times, in three different conditions: a) blind tasting, b) tasting with vision and c) tasting with vision and information about the product. During the blind tasting no information about the product was provided. The tasting with vision is the condition where the participants were merely exposed to the vision of the product (a plastic transparent and closed bag, containing the snack without any logo and/or brand). Finally, in the tasting condition with information, the participants were exposed to the vision of an audiovisual video about the natural composition of the edible coating just before the tasting. At the end of the experiment, participants rated a likert scale where they reported their perception about: pleasantness, the perceived healthiness and their intention to buy the product. During the 3 experimental sessions above described EEG and skin conductance were recorded at 250 Hz by means of *FlexComp infiniti Technology.*

Results

The AW Index shows a positive value in the blind tasting phase, which increases considerably when the tasting is accompanied by the vision of the product. In the third condition, the AW Index, remains positive although slightly lower than the second phase. This results support the claim that information positively affected the experience of tasting and the effect is particularly strong for visual information.

An Approach-withdrawal (AW) index was computed as the difference between the power of the EEG signal in the right and left regions of prefrontal cortex, according to the scientific literature that shows how a greater inhibition of the Alpha band in the left prefrontal cortex predicts a tendency to approach or to appreciate the product. Higher level in this biological AW index means a more important desire and appreciation, while a greater inhibition of the Alpha band in the right region of the same cortical areas predicts a tendency to reject, dislike and avoid the stimulus presented (Sutton and Davidson, 1997; Harmon-Jones, Gable, Peterson, 2010; Price, Peterson, Harmon-Jones, 2012). On the opposite, a greater activation of the left hemisphere, corresponding to a dicreased values in Alpha EEG waves, means a positive emotional reaction and an appreciation of the stimulus, while a lower activation (corresponding to a higher levels in alpha EEG signals) of the left hemisphere is associated to the aversion of de stimulus. This index, called also Approach/Withdraw Index (AW Index), in blind tasting shows positive values, around almost 2 as mean value. In the secondo condition, the same index become more than the double, as it is considerbly increased when the tasting is accompanied by the vision of the product. The signal remains positive also in the third condition although slightly lower than the second condition. Both second and third condition has a significant difference in comparison to the first condition ($p<0.05$, T student test). This result shows that the information provided both by seeing the food product and by the communication about the product is positively affecting the experience of tasting, even if the effect is stronger with the visual information. The index was compared for the whole sample of participants, during the three experimental conditions of tasting. In Tab. 1 the results are showed.

CONDITION	**AW INDEX**
Blind tasting	1.997
View of the product and tasting	4.189
With of the product with communication and tasting	3.918

Tab. 1 – mean AW index values in the three phases of the experiment

The results about skin conductance show a growing signal with the increase of the available information. This trend in different experimental phases is consistent with the results of the study conducted by de Wijk and colleagues (2012) that showed how the vision of food causes a relatively low arousal that grows at the increase of sensory information. The difference between the first and second condition was not significant ($p>0,05$, T student test), while the difference between the first and third condition was significant ($p<0.05$, T student test). The addition of information causes an arousal activation, reflected by the skin conductance, and this can be interpreted as a positive emotional state, according to results obtained from electroencephalographic measurements (AW index) showed above and also reported in the scientific literature. In Tab. 2 are shown the results of skin conductance levels for each experimental conditions.

CONDITION	SKIN CONDUCTANCE
Blind tasting	1.838
View of the product and tasting	2.976
View of the product with communication and tasting	3.172

Tab. 2 – mean Skin Conductance values in the three phases of the experiment

Finally, the participants were asked to evaluate on a Likert scale, ranging from one to five points, the pleasantness of the taste, the perception of healthiness and the willingness to buy after each tasting condition. The condition of tasting with all information was compared with the condition of tasting with visual information only, the evaluation of the participants were higher in the first condition, showing that the information gathered lead a better perception of the food tasting. When the subjects know that the product have genuine characteristics as informed by the communication in third experimental condition, they have an higher intention to buy it. However, the scores did not show significant differences in statistical analyses ($p>0.05$, T student test). In Tab.3 are shown the mean values for the evaluation of the product.

	"View of the product and tasting"	"Wiew of the product with communication and tasting"
Taste	3.95	4.56
Healthiness	3.41	4.27
Willing to buy	3.63	4.39

Tab. 3 – Mean scores of self-report evaluations in the three phases of the experiment

Conclusion

In comparison to the condition of tasting without any information (blind), the subject's evalutation, in terms of AW index, skin conductance levels and self-reports, were more positive when participants were aware about the genuine cheracteristics of the product tasted (third condition, named "*View of the product with communication and tasting")*. Moreover, the perception of healthiness was not the only property affected by the information of genuine features, as also the evalutation of taste was higher too. This experiment points out the importance of communication for food products. It should be noted that there were not significant differences between the second condition (tasting with visual information only) and the third one (visual information + communication), nevertheless there was an increase of activation (in terms of skin conductance activity) and evaluation (in terms of self-reports) consequent to the increase of information about the food product tasted. Last but not least, the AW Index in the third phase showed higher values than in the first condition, and also the skin conductance was significantly higher in the third condition in comparison to the first one (but not between the first and the second one).

References

1. Arcia P., Curutchet A., Costell E., Tarrega A. (2012). Influence of expectations created by label on consumers acceptance of Uruguayan low-fat cheeses. Journal of Sensory Studies, 27, 344–351.
2. Ares, G., Giménez, A., Deliza, R. (2010). Influence of three non-sensory factors on consumer choice of functional yogurts over regular ones. Food Quality and Preference, 21, 361–367.
3. Codeluppi V. (2010), Dalla produzione al consumo. Processi di cambiamento delle società contemporanee, Franco Angeli, Milano.
4. De Wijk R.A., Valesca, K., Verhoeven, R.H.G., Holthuysen, N.T.E., De Graaf, C. (2012). Autonomic nervous system responses on and facial expressions to the sight, smell, and taste of liked and disliked foods. Food Quality and Preference, 26, 196–203.
5. Harmon-Jones, E., Gable, P.A. & Peterson, C.K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. Biological Psychology, 84, 451–462.
6. Itti L., Koch C. (2001), Computational modelling of visual attention, Nature Reviews Neuroscience, 2, 194-203.
7. Mannan, S. K., Kennard, C., Husain, M. (2009), The role of visual salience in Marshall C., Rossman G.B. (1999), eds., Designing Qualitative Research, 3rd ed., Sage, Thousand Oaks.
8. Milosavljevic M., Navalpakkam V., Koch C., Rangel A. (2011), Relative visual saliency differences induce sizable bias in consumer choice. Working Paper, California Institute of Technology.
9. Olivero N., Russo V. (2009), Manuale di psicologia dei consumi. Individuo, società, comunicazione, McGraw-Hill, Milano.
10. Olivero N., Russo V. (2013), Psicologia dei consumi, McGraw-Hill, Milano.
11. Price T.F., Peterson, C.K., Harmon-Jones E. (2012). The emotive neuroscience of embodiment. Motivation and Emotion, 36, 27-37.
12. Russo V. (2015) Neuromarketing, comunicazione e comportamenti di consumo. Principi, strumenti e applicazioni nel food and wine, Franco Angeli, Milano.
13. Siri G. (2001), La psiche del consumo, Franco Angeli, Milano.
14. Sutton S.K., Davidson, R.J. (1997). Prefrontal brain asymmetry: a biological substrate of the behavioral approach and inhibition systems. Psychological Science, 8, 204–210.

Simultaneous fMRI and Eyeblink Conditioning in the Awake Rabbit

Craig Weiss[1], Daniele Procissi[2] and John Disterhoft[1]

[1]Department of Physiology
[2]Department of Radiology,Northwestern University Feinberg School of MedicineChicago, IL 60611

Recent advances in behavioral control and measurement, as well as new techniques of imaging, allow studies in both humans and animals that combine brain imaging as behavior occurs in the conscious, behaving subject. Non-human animal subjects allow the use of interventions that might not be possible in humans and use of such model systems allow for a dissection of the neuronal networks mediating learning and memory.

Functional magnetic resonance imaging (fMRI) provides the opportunity for imaging wide areas of the brain simultaneously with good temporal resolution, but it requires that the subject remain motionless for the duration of the experiment in order to avoid image artifacts. We took advantage of the rabbit's natural tolerance for restraint and utilized MR compatible materials and procedures for conditioning the blink reflex of rabbits while they were positioned inside of the magnet. A "headbolt" was surgically implanted prior to imaging and it was used to secure the head in a standard stereotaxic reference frame. We have utilized both visual stimuli and whisker vibration as conditioning stimuli and have successfully conditioned rabbits to blink without the need for any prolonged habituation to the imaging environment and without using any sedatives.

Our initial investigation revealed that we could evoke blood oxygen level dependent (BOLD) responses to visual stimuli. The responses appeared in the contralateral side visual cortex and lateral geniculate nucleus, in agreement with the known anatomy. Pairing the visual stimulus with a puff of air to the eye induced conditioned blinks to the visual stimulus, as detected with an infrared reflective sensor. The magnitude and percentage of trials with conditioned responses (CRs) increased across training sessions, but not in animals presented with explicitly unpaired stimuli. An analysis of the BOLD response to paired stimuli revealed activation in the cerebellar deep nuclei and deactivation in the cerebellar cortex, consistent with current theories on the cerebellar involvement of eyeblink conditioning.

Based upon the large literature base for the whisker system in rodents, and the ease with which to select the receptive field, i.e. individual whiskers, we stimulated whiskers in the awake immobilized rabbit and detected BOLD responses in the contralateral whisker cortex and ventral posterior medial thalamus. Stimulation of the whiskers proved to be an effective conditioning stimulus. An independent component analysis (ICA) of the intrinsic functional networks of the awake and resting rabbit revealed seven networks similar to those found in humans and rodents, e.g. hippocampal, cerebellar and default mode. The connectivity of these networks was found to change with whisker-signaled trace eyeblink conditioning such that the contralateral somatosensory cortex and ipsilateral deep cerebellar nucleus had greater functional connectivity in the time prior to trials with conditioned responses than during the time prior to trials without a conditioned response.

Although the rabbit is a very good animal model to study functional network activity, we also examined activity at the molecular level by using the MR contrast agent manganese chloride. This compound enters neurons through voltage dependent calcium channels and the manganese remains sequestered for a relatively long time. Subcutaneous injections of manganese chloride were found to increase T1 weighted signal intensity in whisker related areas after a period of six hours; the enhancement was still present when imaged 24 hours later. Thus manganese enhanced magnetic resonance imaging (MEMRI) allows for a molecular analysis of neuronal activity.

The rabbit is thus an excellent experimental subject for using imaging to study neuronal changes that mediate behavior in the awake animal.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Ethics Statement: All experimental procedures were approved by the Northwestern University Institutional Animal Care and Use Committee and performed in accordance with the animal welfare guidelines of the National Institutes of Health.

Combining Spatial Navigation in Virtual Reality with Sub-cellular Resolution Imaging of Neuronal Activity in Mice

M.E.J. Sheffield and D.A. Dombeck

Department of Neurobiology, Northwestern University, Evanston, USA.
m-sheffield@northwestern.edu

The firing discharges of neurons in a brain region called the hippocampus are robustly correlated with an animal's position in its environment. These neurons are called "place cells" and the location in which they fire is called the "place field". Together, these neurons form a cognitive map of the animal's environment, allowing them to navigate accurately from one location to another, much like a GPS system. Traditional methods used to measure place cell firing use electrodes implanted into the hippocampus of rodents while they navigate an enclosed environment (typically ~1 m^2). But these methods are limited in terms of the number of cells that can be simultaneously measured (tens to hundreds), the ability to track the same cells over time (hours to a few days) and the resolution of recording (cannot obtain sub-cellular measurements e.g. from dendrites). This has set limits on our understanding of the neuronal processes underlying spatial navigation. However, these restrictions can be overcome by using a two-photon microscope to measure florescent changes of calcium indicators (a proxy for neuronal firing) in large numbers of cells (thousands), over long periods (weeks to months) and at sub-cellular resolution. A problem with this approach in the context of spatial navigation is that two-photon imaging of the brain requires the neurons to be stable, yet navigation requires the animal to be moving. To overcome this, we surgically attached plates to the heads of mice that could be screwed down to clamp brain motion by restricting head movement (all experiments were approved by the Northwestern University Animal Care and Use Committee). These mice are free to move their limbs to maneuver a spherical treadmill in any direction. Treadmill movement is then translated into movement in a virtual environment displayed around the mice (see Figure 1). Mice are then trained to navigate up and down a virtual linear track for water rewards. During this behavior we imaged large populations of place cells (see Figure 1) over many days and were also able to measure the dendrites and spines (the location on neurons that receive input from other neurons) of place cells [1]. This method has led to new insights into how the brain creates a cognitive map of space, and can be used to study other complex rodent behaviors such as decision making, fear conditioning, motor learning and discrimination tasks. The virtual reality set-up is ideal for gaining precise control over the rodent's visual experience. Coupling this to the ability to measure neuronal activity with two-photon microscopy opens up many doors in the quest to understand the neuronal underpinnings of intricate animal behavior.

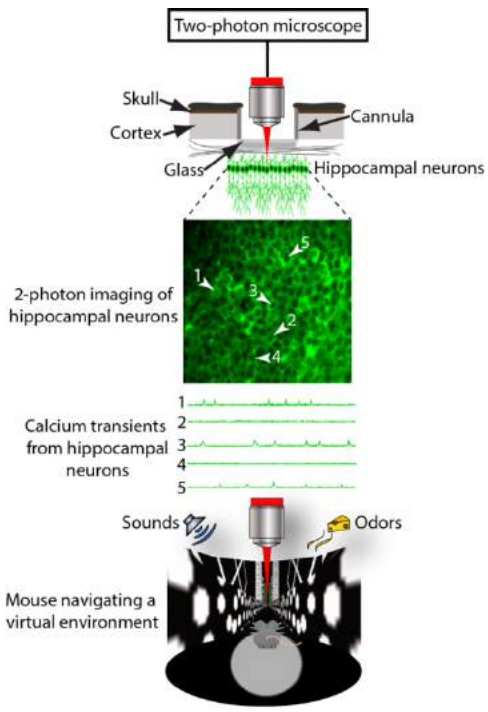


Figure 1. Experimental set-up showing two-photon population imaging of hippocampal neurons in mice during virtual environment navigation.

Reference

1. Sheffield, M.E. & Dombeck, D.A. (2015). Calcium transient prevalence across the dendritic arbour predicts place field properties. Nature 517, 200-204.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Demixing Cellular Signals in Densely-Packed Neuronal Populations

Andrea Giovannucci[1,2], Aleksandra Badura[2], Ben Deverett[2,3], Eftychios Pnevmatikakis[1], Liam Paninski[4] and Samuel S.-H. Wang[2]

[1]Simons Foundation, NYC, NY, United States of America;
[2]Princeton Neuroscience Institute and Department of Molecular Biology, Princeton University, Princeton, NJ
[3]Robert Wood Johnson Medical School, New Brunswick, NJ, USA
[4]Department of Statistics. Columbia University, New York, NY, USA

Abstract

Cerebellar granule cells sit at a point of maximum network divergence in the brain, and monitoring their detailed activity along with behavior would be of great use in understanding neural information processing. However, such monitoring is typically done under anesthesia and/or via aggregated field recordings, since recording granule cells has proven to be extremely difficult due to anatomical and optical features associated with their geometry. Firstly, their density (2600/100 μm3 or xx/100 μm^2), depth (150-300 μm) and highly scattering nature (they act like small lenses) limit optical access for brain imaging and cause high optical and electrical signal overlap. Secondly, fluorescence signals from ascending axons of underlying granule cells sum up creating large synchronized neuropil fluorescence fluctuations. Thirdly, high movement of cerebellar tissue during mouse locomotion induces large motion artifacts. To overcome these difficulties we used a stabilized awake preparation along with custom motion correction software, a cell-type-specific expression of a fast and bright indicator, a ROI and trace selection methods based on matrix factorization. The method we propose allows tracking, in awake behaving mice, the same neuronal populations for days to weeks while tightly monitoring their behavioral state.

In order to overcome optical challenges associated with deep-tissue imaging, we used a high SNR and fast kinetic calcium indicator, GCaMP6f. GCaMP6f is more than twice brighter and over three times faster than GCaMP3[1], which was previously used to image cerebellar granule cells[2][1]. To ensure high levels of indicator expression in GCs we injected AAV1 carrying GCaMP6f under the human synapsin promoter and subsequently waited 3-5 weeks before starting imaging. This approach resulted in a big increase in signal amplitude (30-80% ΔF/F versus previously reported 5-20% ΔF/F) and speed (<170 msec versus >300 msec decay times). We used confocal microscopy of postmortem tissue sections to confirm that granule cells were stained and quantify their expression density 199±20 GrCs per 100 μm2. Our analysis pipeline found 86±29 GrCs (N=4 FOVs in N=4 animals) per 100 μm2 (corresponding to 43% of the expected granule cells).

To mitigate the effect of brain motion the animal preparation was modified from [2] by incorporating skull stabilization. In particular, we cemented the animal skull along all the exposed sutures (lambdoid, sagittal, and coronal sutures) before opening the craniotomy. This removed any degree of freedom for the skull plates relative movement, thus adding stability to the preparation. Although motion artifacts were significantly reduced, they were still present. We removed the residual motion by applying state-of-the-art imaging registration techniques (www.opencv.org template matching) and reduced warping artifacts (due to the overlapping effect of motion and scanning) by dividing the movie into two portions (rostral and caudal) and separately motion correcting them.

The neuropil in the cerebellar GC layer is very dense and compact and the fluorescent signal coming from ascending axons can easily contaminate the somatic one. Previous approaches based on manual ROI selection[2]

[1] Experimental procedures were approved by the Princeton University Institutional Animal Care and Use Committee and performed in accordance with the animal welfare guidelines of the National Institutes of Health. Details of animal preparation were modified from previously published procedures[2].

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

or Principal Component Analysis/Independent Component Analysis[3] failed to separate the two origins of signal types. In order to achieve this spatial separation we used a recent technique based on matrix factorization[4] that has been especially designed for calcium imaging data. This algorithm is specialized in separating signals from overlapping structures by relying on the fact that signals from different structures often have very different time signatures. This algorithm removes activity from diffused sources, a pattern consistent with ascending axons passing through neuropil, as well as de-mixing the contributions of partially overlapping cells.

To monitor fine movements of whisker, eyelid and forelimbs of a mouse with ~15 ms resolution we used an infrared camera followed by image recognition and extraction. We use these methods to observe responses to sensory and motor events within overlapping populations of GCs. As a case study, we imaged the activity of granule cell populations over multiple days of eyeblink conditioning, a type of learning that requires the cerebellum. During the first few days of conditioning, we found groups of granule cells that responded to neutral visual and somatosensory stimuli, as well as the unconditional periorbital airpuff used to train the associative response. Subsequently, the same granule cells gradually acquired a conditional response that grew with training and whose timing approximately matched that of the learned movement. The activity of individual granule cells co-varied strongly and accurately represented behavioral performance on a trial-by-trial basis, forming a code that was highly redundant. Thus, information about learned associations is available at the input stage of the cerebellar cortex, providing the type of online motor feedback that is necessary in modern models of cerebellar function.

References

1. Badura, A., et al., Fast calcium sensor proteins for monitoring neural activity. Neurophotonics, 2014. 1(2): p. 025008.
2. Ozden, I., et al., Widespread state-dependent shifts in cerebellar activity in locomoting mice. PLoS One, 2012. 7(8): p. e42650.
3. Mukamel, E.A., A. Nimmerjahn, and M.J. Schnitzer, Automated analysis of cellular signals from large-scale calcium imaging data. Neuron, 2009. 63(6): p. 747-60.
4. Pnevmatikakis, E.A., et al., Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron, 2016. 89(2): p. 285-99.

Assessing Implicit and Explicit Memory in Humans During Functional Brain Imaging

Fred J. Helmstetter

Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI USA.
fjh@uwm.edu

In even the simplest behavioral experiments, human subjects and laboratory animals learn and remember multiple aspects of what they experience. Basic associative learning procedures like Pavlovian conditioning are no exception, and human participants exposed to programmed relationships between simple stimuli normally become consciously aware of these relationships and can describe them. However, they also can show implicit evidence of learning that is robust but not accessible to conscious awareness. The similarities and differences between explicit and implicit processes have been important in evaluating various theoretical approaches to understanding learning and memory. The distinction is also important for understanding brain systems that subserve specific behavioral capabilities.

This symposium presentation will provide an overview of several of our recent studies on aversive Pavlovian (fear) conditioning with human subjects using functional magnetic resonance imaging (FMRI) and magnetoencephalography (MEG). In all of our experiments, we relate participants' explicit knowledge of stimulus relationships with measures of implicit memory and with measurement of brain activity in real time during the learning process or memory retrieval trials after training. I will discuss the relative mertits of various approaches to characterizing implicit fear responses including the use of galvanic skin response, pupil dilation, and tracking covert eye movements. A number of behavioral factors including task complexity, emotional salience, and memory consolidation will also be considered. Much of the work I will present uses a training protocol in which a breif shock (UCS) follows a visual stimulus (CS) after an empty interval of time (trace conditining) which, unlike standard "delay" conditioning, requires processing in prefrontal cortex and medial temporal lobe structures. The goal is to provide an understanding of how multiple behavioral dependent variables within a single experiment can be related to functional brain imaging data both during task performance and in quantifying resting state functional connectivity.

Ethics statement: All experiments described have been approved by the Institutional Review Board for the protection of human subjects at the University of Wisconsin-Milwaukee and at the Medical College of Wisconsin.

Reference

1. Hopkins, L.S., Schultz, D.H., Hannula, D.E. & Helmstetter, F.J. (2015) Eye movements index implicit memory expression in fear conditioning. PLoS ONE, 10: e0141949. doi: 10.1371/journal.pone.0141949
2. Balderston, N.L., Schultz, D.H., Hopkins, L. & Helmstetter, F.J. (2015) Functionally distinct amygdala subregions identified using DTI and high resolution fMRI. Social, Cognitive and Affective Neuroscience, 10: 1615-1622, doi: 10.1093/scan/nsv055
3. Balderston, N.L., Schultz, D.H. Baillet, S. & Helmstetter, F.J. (2014) Rapid amygdala responses during trace fear conditioning without awareness. PLoS ONE. doi:10.1371/journal.pone.0096803
4. Schultz, D.H., Balderston, N.L., & Helmstetter, F.J. (2012) Resting-state connectivity of the amygdala is altered following Pavlovian fear conditioning, Frontiers in Human Neuroscience,.6:242 doi:10.3389/fnhum.2012.00242
5. Gilmartin, M.R., Balderston, N.L. & Helmstetter, F.J. (2014) Prefrontal cortical regulation of fear learning. Trends in Neuroscience, 37:455-464, doi: 10.1016/j.tins.2014.05.004.

Functional Monitoring of Home Dwelling Alzheimer's Patients

Katrin Schenk , George Netscher , Brad Zylstra , Julian Jacquemot , Steve Bonasera
Evan Goulding

Randolph College, United States of America

We will discuss the development of a system, the Functional Monitoring System, that uses ubiquitous computing devices (e.g. cellphones, bluetooth sensors, and smartwatches) to measure how individuals pattern and localize their moment-to-moment activities in the home and in the community. This data is used to assess functional status in a continuous, near-real time, inexpensive, easy-to-use, secure, scalable, precise, and validated manner [1]. This kind of data can be classified into time series representations of important behaviors, and can provide important prognostic and therapeutic information for many community dwelling people, including elderly individuals at risk for functional loss and persons with neurodegenerative diseases such as Alzheimer's Disease.

We classify the data to determine activity levels and location in the home (homespace) and community (lifespace). These metrics provide a nuanced and unbiased picture of an individual's current status predictive of health status outcomes [2-4]. Dementia patients may not be able to identify or communicate changes in health status, but identification of changes in activity, homespace [5] and lifespace, [5,6] can allow health care providers to quickly intervene and potentially prevent a major crisis. Functional Monitoring can also identify insidious decreases in activity, homespace, and lifespace characteristic of functional decline, allowing health care providers to initiate treatment plans aimed at maintaining independence at home. The main burden for participants is keeping the phone charged and on their person, and wearing a watch. No other effort is required. This system is currently being tested in a large, Medicare funded cilinical trial in collaboration with the University of California, San Francisco and the University of Nebraska, Medical Center.

The protocol for continuous functional assessment in this study is as follows. Patients (or their caregiver) receive a Moto-G cell phone running the Android operating system. They will be shown how to charge the phone and turn it on. Staff will visit their home for equipment deployment. Simple, inexpensive environmental sensors (Estimotes (www.estimote.com)) will be placed in all home rooms. Subjects wear a "smart" wristwatch (Sony SW3) which communicates with the Estimotes and records the signal strength of each beacon as the patient moves around the house. This data is continuously uploaded to the Moto-G phone. The phone then uploads the activity, GPS and indoor beacon signal strength data once a day to our secure servers.

Once data is sent to the study server, it is decrypted and placed in a database. Functional indices are calculated and compared to prior examples of that subject's function. Acute changes (e.g., increased or decreased activity at a given circadian time point, sudden changes in homespace such as increased time in bathroom or bedroom) will generate an alert on a web based "Dashboard" (written in Salesforce) alert that will be monitored by clinical personnel. Similarly, changes in functional indices occurring over longer (e.g., 2-week) time courses (e.g., decreased lifespace, small changes in gait speed or step count) will generate an alert indicating more insidious functional loss.

1. Schenk AK*, et al.* (2011) Cellular telephones measure activity and lifespace in community-dwelling adults: proof of principle. J Am Geriatr Soc, 59(2), 345-352.
2. Boyle PA*, et al.* (2010) Association between life space and risk of mortality in advanced age. J Am Geriatr Soc, 58(10), 1925-1930.
3. Xue QL*, et al.* (2008) Life-space constriction, development of frailty, and the competing risk of mortality: the Women's Health And Aging Study I. Am J Epidemiol, 167(2), 240-248.
4. Blair SN*, et al.* (1995) Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. JAMA, 273(14), 1093-1098.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

5. Campbell I. H., *et al.* (2011) Measuring changes in activity patterns during a norovirus epidemic at a retirement community. Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE; 2011: IEEE: 6793-6796.
6. Brown CJ, *et al.* (2009) Trajectories of life-space mobility after hospitalization. Ann Intern Med, 150(6), 372-378.

LiveWell: A Smartphone Application to Enhance Understanding and Treatment of Bipolar Disorder

Evan Goulding[1], Cynthia Dopke[1], Clair Martin[1], Yong Jun Kwon[2], Tania Michaels[1], Mark Begale[1], Chris Karr[1], Katrin Schenk[2], Stephen Bonasera[3], David Mohr[1]

[1]Department of Psychiatry, Northwestern University, Chicago, USA.
e-goulding@fsm.northwestern.edu
[2]Department of Physics, Randolph College, Lynchburg, USA.
kschenk@randolphcollege.edu
[3]Department of Medicine, University of Nebraska Medical Center, Omaha, USA.
sbonasera@unmc.edu

Bipolar disorder is a chronic mental illness with high levels of morbidity and mortality [1, 2]. Even when receiving current pharmacological treatment, individuals with bipolar disorder often suffer multiple episodes, long episode durations, and inter-episode subsyndromal symptoms [1, 3]. Combining psychotherapeutic interventions for bipolar disorder with pharmacological treatment improves clinical outcomes [4-13]. However, these psychotherapeutic interventions are resource intensive, and few patients receive treatment with evidence-based therapies [4-8, 14, 15]. Standard psychosocial interventions also do not provide real-time assessment and feedback which could be used to optimize treatment outcomes. In addition, very little data exists that specifies what behavioral targets mediate the benefits of bipolar disorder psychotherapeutic interventions [7, 9, 11, 16]. These deficiencies limit the impact of psychosocial interventions for bipolar disorder and hamper our ability to improve treatment.

Behavioral intervention technologies (BITs), including smartphone interventions, can increase access to psychosocial interventions with demonstrated efficacy thereby enhancing treatment. BITs also provide a platform for measuring behaviors relevant to bipolar disorder phenotypes, treatment course, and intervention mechanisms [17]. Early development and testing of web-based and mobile interventions for bipolar disorder is underway [18]. Several internet interventions have been found to be feasible and highly acceptable methods and adding human support improves adherence with use [18, 19]. Similar to standard psychosocial interventions, simply providing weekly psychoeducation via the internet has limited effects [19-21]. However, incorporating active use of cognitive behavioaral therapy tools along with psychoeducation results in a reduction in depressive symptoms [22]. The design and development of mobile phone interventions for bipolar disorder has been reported but only limited early outcome data is available [18]. These efforts have focused on: behavioral monitoring reviewed weekly by mental health providers; face-to-face (F2F) therapy augmented by self-monitoring and personalized short messaging feedback [18, 23]

We are developing a smartphone-based BIT, LiveWell, that aims to provide real-time personalized feedback based on self-report and behavioral sensor data to assist individuals with managing bipolar disorder in collaboration with their psychiatrist. The intervention utilizes a coach operating within the supportive accountabilty model as means to improve outcomes as we convert a F2F therapy to a BIT [24]. Here, we will describe the design and initial data collection and analysis efforts for LiveWell. We will focus on two main areas of behavioral measurement and analysis.

First, we will address how the content and architecture of a BIT, like LiveWell, can be developed to facilitate enhanced understanding of psychosocial intervention mechanisms. This is an important area for improving behavioral measurement and analysis because the targets of psychosocial interventions for bipolar disorder are not well understood and very little data on moderators and mediators exists [7, 9, 11, 16]. BITs lend themselves to studying how psychosocial interventions work as they provide a means to track how users interact with the intervention. The role of explicitly defining and labeling performance objectives and underlying determinants targeted for change along with behavioral change techniques delivered will be discussed as a means for dissecting intervention effects and paving the way for intervention tailoring and improvement.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Next, we will address the role of utilizing LiveWell to measure and analyze behavioral patterns that are highly relevant to treatment course and potential endophenotypes in bipolar disorder. Sleep, activity, and circadian rhythm disturbances are prominent features of subsyndromal, prodromal, and episodic disturbances in bipolar disorder [25-37]. LiveWell automates long-term collection of self-report, activity/sleep, location, and social interaction data using Purple Robot (open-source Android application) to collect smartphone sensor and peripherals data and to transmit all data [38]. Issues related to continuous collection of data from phone sensors and wrist-worn actimeters and automated quality control and data processing will be adressed. This will be followed by a discussion of initial data analysis techniques used to explore temporal patterns of change in activity and location and their relationship to clinical status and course. By analyzing self-report and behavioral data along with application usage data and meta-data, we hope to improve intervention impact, better understand phenotypes and course, and gain insight into the mechanisms that mediate the effectiveness of psychosocial interventions for bipolar disorder.

This study was approved by the Institutional Review Board of Northwestern University and was registered in ClincalTrials.gov (Identifier NCT02405117).

Reference

1. Eaton, W.W., S.S. Martins, G. Nestadt, et al. (2008). The burden of mental disorders. Epidemiol Rev. 30, 1-14.
2. Dilsaver, S.C. (2011). An estimate of the minimum economic burden of bipolar I and II disorders in the United States: 2009. J Affect Disord. 129(1-3), 79-83.
3. Judd, L.L., H.S. Akiskal, P.J. Schettler, et al. (2002). The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry. 59(6), 530-7.
4. Crowe, M., L. Whitehead, L. Wilson, et al. (2010). Disorder-specific psychosocial interventions for bipolar disorder--a systematic review of the evidence for mental health nursing practice. Int J Nurs Stud. 47(7), 896-908.
5. Miklowitz, D.J. (2008). Adjunctive psychotherapy for bipolar disorder: state of the evidence. Am J Psychiatry. 165(11), 1408-19.
6. Miklowitz, D.J., G.M. Goodwin, M.S. Bauer, et al. (2008). Common and specific elements of psychosocial treatments for bipolar disorder: a survey of clinicians participating in randomized trials. J Psychiatr Pract. 14(2), 77-85.
7. Miklowitz, D.J. and J. Scott (2009). Psychosocial treatments for bipolar disorder: cost-effectiveness, mediating mechanisms, and future directions. Bipolar Disord. 11 Suppl 2, 110-22.
8. Schottle, D., C.G. Huber, T. Bock, et al. (2011). Psychotherapy for bipolar disorder: a review of the most recent studies. Curr Opin Psychiatry. 24(6), 549-55.
9. Geddes, J.R. and D.J. Miklowitz (2013). Treatment of bipolar disorder. Lancet. 381(9878), 1672-82.
10. Lam, D.H., R. Burbeck, K. Wright, et al. (2009). Psychological therapies in bipolar disorder: the effect of illness history on relapse prevention - a systematic review. Bipolar Disord. 11(5), 474-82.
11. Scott, J., F. Colom and E. Vieta (2007). A meta-analysis of relapse rates with adjunctive psychological therapies compared to usual psychiatric treatment for bipolar disorders. Int J Neuropsychopharmacol. 10(1), 123-9.
12. Bond, K. and I.M. Anderson (2015). Psychoeducation for relapse prevention in bipolar disorder: a systematic review of efficacy in randomized controlled trials. Bipolar Disord. 17(4), 349-62.
13. Gregory, V.L., Jr. (2010). Cognitive-behavioral therapy for depression in bipolar disorder: a meta-analysis. J Evid Based Soc Work. 7(4), 269-79.
14. Busch, A.B., D. Ling, R.G. Frank, et al. (2007). Changes in the quality of care for bipolar I disorder during the 1990s. Psychiatr Serv. 58(1), 27-33.
15. Wang, P.S., M. Lane, M. Olfson, et al. (2005). Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. Arch Gen Psychiatry. 62(6), 629-40.
16. Beynon, S., K. Soares-Weiser, N. Woolacott, et al. (2008). Psychosocial interventions for the prevention of relapse in bipolar disorder: systematic review of controlled trials. Br J Psychiatry. 192(1), 5-11.
17. Mohr, D.C., M.N. Burns, S.M. Schueller, et al. (2013). Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry. 35(4), 332-8.
18. Hidalgo-Mazzei, D., A. Mateu, M. Reinares, et al. (2015). Internet-based psychological interventions for bipolar disorder: Review of the present and insights into the future. J Affect Disord. 188, 1-13.

19. Proudfoot, J., G. Parker, V. Manicavasagar, et al. (2012). Effects of adjunctive peer support on perceptions of illness control and understanding in an online psychoeducation program for bipolar disorder: a randomised controlled trial. J Affect Disord. 142(1-3), 98-105.
20. Barnes, C.W., D. Hadzi-Pavlovic, K. Wilhelm, et al. (2015). A web-based preventive intervention program for bipolar disorder: outcome of a 12-months randomized controlled trial. J Affect Disord. 174, 485-92.
21. Lehner, R.K., Dopke, C.A., Cohen, K., Edstrom, K., Maslar, M., Slagg, N.B., Yohanna, D. (2007). Outpatient Treatment Adherence and Serious Mental Illness: A Review of Interventions. American Journal of Psychiatric Rehabilitation. 10, 245-274.
22. Lauder, S., A. Chester, D. Castle, et al. (2015). A randomized head to head trial of MoodSwings.net.au: an Internet based self-help program for bipolar disorder. J Affect Disord. 171, 13-21.
23. Depp, C.A., J. Ceglowski, V.C. Wang, et al. (2015). Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. J Affect Disord. 174, 23-30.
24. Mohr, D.C., P. Cuijpers and K. Lehman (2011). Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res. 13(1), e30.
25. Harvey, A.G. (2008). Sleep and circadian rhythms in bipolar disorder: seeking synchrony, harmony, and regulation. Am J Psychiatry. 165(7), 820-9.
26. Gruber, J., D.J. Miklowitz, A.G. Harvey, et al. (2011). Sleep matters: sleep functioning and course of illness in bipolar disorder. J Affect Disord. 134(1-3), 416-20.
27. Bauer, M., P. Grof, N. Rasgon, et al. (2006). Temporal relation between sleep and mood in patients with bipolar disorder. Bipolar Disord. 8(2), 160-7.
28. Bauer, M., G. Juckel, C.U. Correll, et al. (2008). Diagnosis and treatment in the early illness phase of bipolar disorders. Eur Arch Psychiatry Clin Neurosci. 258 Suppl 5, 50-4.
29. Jackson, A., J. Cavanagh and J. Scott (2003). A systematic review of manic and depressive prodromes. J Affect Disord. 74(3), 209-17.
30. Harvey, A.G., D.A. Schmidt, A. Scarna, et al. (2005). Sleep-related functioning in euthymic patients with bipolar disorder, patients with insomnia, and subjects without sleep problems. Am J Psychiatry. 162(1), 50-7.
31. Murray, G. and A. Harvey (2010). Circadian rhythms and sleep in bipolar disorder. Bipolar Disord. 12(5), 459-72.
32. Jones, S.H. (2001). Circadian rhythms, multilevel models of emotion and bipolar disorder--an initial step towards integration? Clin Psychol Rev. 21(8), 1193-209.
33. Johnson, S.L. (2005). Life events in bipolar disorder: towards more specific models. Clin Psychol Rev. 25(8), 1008-27.
34. Ehlers, C.L., E. Frank and D.J. Kupfer (1988). Social Zeitgebers and Biological Rhythms - A Unified Approach to Understanding the Etiology of Depression. Archives of General Psychiatry. 45(10), 948-952.
35. Frank, E. (2007). Interpersonal and social rhythm therapy: a means of improving depression and preventing relapse in bipolar disorder. J Clin Psychol. 63(5), 463-73.
36. McCarthy, M.J., H. Wei, Z. Marnoy, et al. (2013). Genetic and clinical factors predict lithium's effects on PER2 gene expression rhythms in cells from bipolar disorder patients. Translational Psychiatry. 3.
37. Yang, S., H.P. Van Dongen, K. Wang, et al. (2009). Assessment of circadian function in fibroblasts of patients with bipolar disorder. Mol Psychiatry. 14(2), 143-55.
38. Schueller, S.M., M. Begale, F.J. Penedo, et al. (2014). Purple: a modular system for developing and deploying behavioral intervention technologies. J Med Internet Res. 16(7), e181.

High-Throughput System for Mouse Home Cage Behavioral Phenotyping

Evan Goulding[1], Yong Jun Kwon[2], Katrin Schenk[2], Stephen Bonasera[3] and Laurence Tecott[4]

1 Department of Psychiatry, Northwestern University, Chicago, USA.
e-goulding@fsm.northwestern.edu
2 Department of Physics, Randolph College, Lynchburg, USA.
kschenk@randolphcollege.edu
3 Department of Medicine, University of Nebraska Medical Center, Omaha, USA.
sbonasera@unmc.edu
4 Department of Psychiatry, University of California San Francisco, San Francisco, USA.
larrytecott@ucsf.edu

Animals in natural environments exhibit spatial and temporal patterns of behavior that are shaped both by the environment and genetics [1-3]. With the advent and increasing use of smartphones and wearable peripheral devices, there has been a rapid expansion in the availability of data capturing daily patterns of human behavior [4-6]. This type of human data has the potential to provide novel ways to understand and intervene in problems related to health maintenance and mental illness [7, 8]. To facilitate the capacity to process and analyze data from humans going about their daily life, it may be useful to have parallel processing and analysis algorithms that can be utilized in more controlled settings [9, 10]. Although several sophisticated approaches for automated behavioral data collection and home cage monitoring exist, they do not primarily focus on algorithms that quantitatively capture the rich temporal and spatial structure of diverse behaviors that occur over multiple time scales in both mice and humans [11-25]. Here, we will discuss methods for processing and analyzing data from continuous home cage behavioral monitoring of ingestion and activity in 16 inbred mouse strains. Techniques used to classify the behavior of the mice into active states characterized by bouts of feeding, drinking, and locomotion and inactive states characterized by prolonged immobility at the nest will be described [26]. The use of these behavioral features to examine genetic differences in some of these basic underlying units of behavioral regulation will then be explored. In addition, methods for analyzing and clustering higher levels of organization, such as temporal patterns of active states, will be discussed. For example, clustering of mouse inbred strains based on the temporal patterns of their active states reveals that many of the inbred strains can be tightly grouped suggesting that complex patterns of daily behavior may be heritable. Some initial examples of porting analysis algorithms from mice to humans will also be briefly reviewed to support the proposal that quantitatively assessing similar behavioral patterns in mice and humans, such as activity

Data reported in this presenation derives from studies approved by the Institutional Review Boards of Northwestern University and the University of Nebraska Medical Center and from the Institutional Animal Care and Use Committees of Northwestern University, University of Nebraska Medical Center, and the University of California, San Francisco.

Reference

1. McFarland, D.J. (1977). Decision Making in Animals. Nature. 269(5623), 15-21.
2. Oatley, K. (1970). Brain mechanisms and motivation. Nature. 225(5235), 797-801.
3. Prescott, T.J., P. Redgrave and K. Gurney (1999). Layered control architectures in robots and vertebrates. Adaptive Behavior. 7(1), 99-127.
4. Blondel, V.D., A. Decuyper and G. Krings (2015). A survey of results on mobile phone datasets analysis. Epj Data Science. 4(1).
5. Schenk, A.K., B.C. Witbrodt, C.A. Hoarty, et al. (2011). Cellular telephones measure activity and lifespace in community-dwelling adults: proof of principle. J Am Geriatr Soc. 59(2), 345-52.
6. Saeb, S., M. Zhang, C.J. Karr, et al. (2015). Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res. 17(7), e175.
7. Hidalgo-Mazzei, D., A. Mateu, M. Reinares, et al. (2015). Internet-based psychological interventions for bipolar disorder: Review of the present and insights into the future. J Affect Disord. 188, 1-13.
8. Mohr, D.C., M.N. Burns, S.M. Schueller, et al. (2013). Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry. 35(4), 332-8.

9. Henry, B.L., A. Minassian, J.W. Young, et al. (2010). Cross-species assessments of motor and exploratory behavior related to bipolar disorder. Neurosci Biobehav Rev. 34(8), 1296-306.
10. van Enkhuizen, J., M.A. Geyer, A. Minassian, et al. (2015). Investigating the underlying mechanisms of aberrant behaviors in bipolar disorder from patients to models: Rodent and human studies. Neurosci Biobehav Rev. 58, 4-18.
11. Casadesus, G., B. Shukitt-Hale and J.A. Joseph (2001). Automated measurement of age-related changes in the locomotor response to environmental novelty and home-cage activity. Mech Ageing Dev. 122(15), 1887-97.
12. Gannon, K.S., J.C. Smith, R. Henderson, et al. (1992). A system for studying the microstructure of ingestive behavior in mice. Physiol Behav. 51(3), 515-21.
13. Chen, S.A., L.E. O'Dell, M.E. Hoefer, et al. (2006). Unlimited access to heroin self-administration: independent motivational markers of opiate dependence. Neuropsychopharmacology. 31(12), 2692-707.
14. de Visser, L., R. van den Bos, W.W. Kuurman, et al. (2006). Novel approach to the behavioural characterization of inbred mice: automated home cage observations. Genes Brain Behav. 5(6), 458-66.
15. Millecamps, M., D. Jourdan, S. Leger, et al. (2005). Circadian pattern of spontaneous behavior in monarthritic rats: a novel global approach to evaluation of chronic pain and treatment effectiveness. Arthritis Rheum. 52(11), 3470-8.
16. Steele, A.D., W.S. Jackson, O.D. King, et al. (2007). The power of automated high-resolution behavior analysis revealed by its application to mouse models of Huntington's and prion diseases. Proc Natl Acad Sci U S A. 104(6), 1983-8.
17. Tamborini, P., H. Sigg and G. Zbinden (1989). Quantitative analysis of rat activity in the home cage by infrared monitoring. Application to the acute toxicity testing of acetanilide and phenylmercuric acetate. Arch Toxicol. 63(2), 85-96.
18. Tang, X. and L.D. Sanford (2005). Home cage activity and activity-based measures of anxiety in 129P3/J, 129X1/SvJ and C57BL/6J mice. Physiol Behav. 84(1), 105-15.
19. Zorrilla, E.P., K. Inoue, E.M. Fekete, et al. (2005). Measuring meals: structure of prandial food and water intake of rats. Am J Physiol Regul Integr Comp Physiol. 288(6), R1450-67.
20. Bonasera, S.J., A.K. Schenk, E.J. Luxenberg, et al. (2008). A novel method for automatic quantification of psychostimulant-evoked route-tracing stereotypy: application to Mus musculus. Psychopharmacology. 196(4), 591-602.
21. Brunner, D., E. Nestler and E. Leahy (2002). In need of high-throughput behavioral systems. Drug Discovery Today. 7(18), S107-S112.
22. Donohue, K.D., D.C. Medonza, E.R. Crane, et al. (2008). Assessment of a non-invasive high-throughput classifier for behaviours associated with sleep and wake in mice. Biomedical Engineering Online. 7.
23. Galsworthy, M.J., I. Amrein, P.A. Kuptsov, et al. (2005). A comparison of wild-caught wood mice and bank voles in the Intellicage: assessing exploration, daily activity patterns and place learning paradigms. Behavioural Brain Research. 157(2), 211-217.
24. Kas, M.J.H., A.D. Malsen, B. Olivier, et al. (2008). Differential genetic regulation of motor activity and anxiety-related behaviors in mice using an automated home cage task. Behavioral Neuroscience. 122(4), 769-776.
25. Risbrough, V.B., V.L. Masten, S. Caldwell, et al. (2006). Differential contributions of dopamine D-1, D-2, and D-3 receptors to MDMA-induced effects on locomotor behavior patterns in mice. Neuropsychopharmacology. 31(11), 2349-2358.
26. 26. Goulding, E.H., A.K. Schenk, P. Juneja, et al. (2008). A robust automated system elucidates mouse home cage behavioral structure. Proc Natl Acad Sci U S A. 105(52), 20575-82.

Systematic PhenoTyper Data Analysis Using AHCODA Analysis Software and AHCODA-DB Data Repository

B. Koopmans[1], A.B. Smit[2], M. Verhage[3] and M. Loos[1]

[1]Sylics (Synaptologics BV), Amsterdam
[2]Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research (CNCR), Neuroscience Campus Amsterdam, VU University, Amsterdam
[3]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research (CNCR), Neuroscience Campus Amsterdam, VU University, Amsterdam

Mouse models play an important role in studying brain disorders and in developing/testing new therapeutic strategies. Characterizing mutant mouse lines and therapeutic strategies in standard behavioural experiments provides a great insight in a large part of the behavioural spectrum and understanding the underlying mechanisms of these disorders and therapeutic strategies. However, conducting these experiments is time consuming and labour intensive. Moreover, extensive human interference with mouse behaviour and human-centred planning of experiments, increases stress and variation in behaviour complicating the systematic comparison of mouse mutants and treatments [1].

In the Neuro-Bsik Mouse Phenomics consortium, an automated home-cage screening platform for testing novel mutant mouse lines and therapeutic strategies was developed. It consists of an automated home-cage (PhenoTyper™ model 3000) with a camera mounted in the top-unit to track the animals using EthoVision software (PhenoTyper model 3000 & EthoVision HTP 2.1.2.0, based on EthoVision XT 4.1 and EthoVision XT 11, Noldus Information Technology, Wageningen, The Netherlands). In order to analyse the wealth of data generated with this automated platform, we created the cloud-based analysis software package AHCODA (Synaptologics BV, Amsterdam, The Netherlands). This software package uses raw frame-by-frame tracking data from EthoVision and generates parameters specific for each experiment conducted. Together, this combination of an automated home-cage, video tracking and cloud-based data analyses allows systematic testing of many mice at the same time, at different test sites, in a shorter time window and without human interference to explore known and novel phenotypes and treatment effects. Due to the continuous development of behavioural testing protocols, this platform can be used to explore different domains of mouse behaviour, such as spontaneous behaviour [2], avoidance learning [3], discrimination- and reversal learning (Remmelink *et al.*, in prep.) and anxiety [4]. For example, the spontaneous behaviour protocol and data analysis generates a set of 115 parameters, of which 105 showed highly significant strain differences among 11 common inbred strains [2].

All individual data points generated by AHCODA for currently available testing protocols are stored in a central database (AHCODA-DB), and currently contains >10,000 mouse profiles. The publicly accessible data contained in this database (published data and several datasets generated with public funding) can be accessed using the AHCODA-DB website (Koopmans *et al.*, in prep., https://public.sylics.com) as open access data. In addition, the AHCODA-DB website also integrates reference data obtained with conventional behavioural tests of several publicly funded projects. The analysis tools integrated in the AHCODA-DB website can be used for data mining of automated home-cage data as well as these conventional behavioural tests, and customised statistical analysis of different (mutant) mouse lines and pharmacological interventions. Using this AHCODA-DB website researchers can systematically compare mutant mouse lines and pharmacological interventions across hundreds of behavioural parameters, potentially providing new insights in underlying mechanisms.

References

1. Crabbe, J.C., Wahlsten, D. and Dudek, B.C. (1999) Genetics of mouse behavior: interactions with laboratory environment. *Science (New York, NY)*, 284, 1670–2. http://dx.doi.org/10.1126/science.284.5420.1670
2. Loos, M., Koopmans, B., Aarts, E., Maroteaux, G., van der Sluis, S., Verhage, M. et al. (2014) Sheltering Behavior and Locomotor Activity in 11 Genetically Diverse Common Inbred Mouse Strains Using Home-Cage Monitoring. PLoS ONE, 9, e108563. http://dx.doi.org/10.1371/journal.pone.0108563

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

3. Maroteaux, G., Loos, M., van der Sluis, S., Koopmans, B., Aarts, E., van Gassen, K. et al. (2012) High-throughput phenotyping of avoidance learning in mice discriminates different genotypes and identifies a novel gene. Genes, Brain, and Behavior, 11, 772–84. http://dx.doi.org/10.1111/j.1601-183X.2012.00820.x
4. Aarts, E., Maroteaux, G., Loos, M., Koopmans, B., Kovačević, J., Smit, A.B. et al. (2015) The light spot test: measuring anxiety in mice in an automated home-cage environment. Behavioural Brain Research, 294, 123–30. http://dx.doi.org/10.1016/j.bbr.2015.06.011

Assessing Discrimination- and Reversal Learning in the PhenoTyper within 4 Days: Specific Impairments in an Alzheimer's Disease Model

E. Remmelink[1,2,3], A.B. Smit[2], M. Verhage[3] and M. Loos[1]

[1]Sylics (Synaptologics BV), Amsterdam
[2]Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research (CNCR), Neuroscience Campus Amsterdam, VU University, Amsterdam
[3]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research (CNCR), Neuroscience Campus Amsterdam, VU University, Amsterdam

Background

In several neurological and psychiatric disorders, different aspects of cognition are affected, e.g learning, working memory, attention, or cognitive flexibility. Measuring cognition in mice is one important approach in understanding the mechanisms underlying cognitive deficits in these disorders and for the development of novel treatment options. The majority of currently available behavioral tests measuring cognitive functions are operant tasks, which require extended training periods and scheduled food deprivation to motivate mice to perform. Here, we describe a novel 4-day continuously operating task measuring discrimination- and reversal learning (CognitionWall DL/RL task) in an automated home-cage that largely eliminates these limitations [1] and we demonstrate its sensitivity and reproducibility to detect learning deficits in a mouse model for Alzheimer's disease.

Methods

Home-cage and task protocol

The task was implemented in an automated home-cage (PhenoTyper model 3000, Noldus Information Technology, Wageningen, The Netherlands), in which behavior was video tracked by a camera mounted in the top. Hardware actions were triggered by the position of the mouse (EthoVision XT 11, Noldus Information Technology, Wageningen, The Netherlands). The cage was equipped with a shelter, a feeding station, and spouts of a water bottle and a reward dispenser.

After an initial three day habituation period to the home-cage during which spontaneous behavior was tracked and analyzed [2], a wall with three holes (CognitionWall) was placed in front of the pellet dispenser spout. For two days, mice had to learn to earn their food (Dustless Precision Pellets, 14 mg, Bio-Serve, Frenchtown, NJ, USA) by going through the left hole in the wall (Discrimination learning (DL)). During this period the dispenser distributed 1 reward for every 5 times the mouse went through this correct left hole (FR5 schedule). The middle and right hole were the incorrect holes and passing through these holes did not have any consequences. During the subsequent 2 days, the correct hole was switched to the right hole (Reversal learning (RL)).

Data analysis

The total number of entries needed to reach a criterion of 80% correct, computed as a moving window with window size 30 (i.e. 24 correct entries of the 30 last entries), was used as the main measure of learning during DL and RL. In addition to this measure, ~100 other measures were extracted that represent e.g. learning, perseverative behavior and general activity.

All data analyses were performed in R and were automatically executed when raw data from Ethovision was added to our AHCODA-DB data repository (Koopmans et al., in prep., https://public.sylics.com).

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Mice

Reference performance of adult C57BL/6J mice was assessed to determine the suitability of this task to study learning and cognitive flexibility within 4 days. Discrimination learning in APPswe/PS1dE9 mice, a common mouse model of Alzheimer's disease that develops amyloid plaques around 6 months of age, was assessed to investigate the sensitivity of the task to detect learning deficits. Experiments were carried out in accordance with the European Communities Council Directive of 24 November 1986 (86/609/EEC), and with approval of the Animal Experiments Committee of the VU University (Protocols SYN1302 and NBC008).

Results

C57BL/6J mice attained the performance criterion of 80% correct within 1 day. Reversal learning requires suppression of the initially acquired response, while learning a new, competing, rule [3]. As expected, mice took significantly longer to attain the performance criterion during the reversal phase, compared to the initial learning phase. Nonetheless, all individuals attained the criterion within the two available reversal learning days. Weight loss in both groups of C57BL/6J mice was 3% after the 4 days of the task.

Several independent cohorts of APPswe/PS1dE9 mice required substantially more entries to reach the 80% performance level during DL compared with their WT littermate controls. Interestingly, this impairment in learning was evident within the first few hours of the task.

Conclusions

We have developed a 4-day protocol for discrimination- and reversal learning, which operates without human intervention. The task was sensitive to detect learning deficits in a mouse model of Alzheimer's disease. Since the learning deficit in APP/PS1 mice was evident within the first few hours of the task, this allows testing of acute (single dose) interventions administered prior to cognitive testing (e.g. cognitive enhancers). Hence, due to its short duration and the absence of scheduled food deprivation and concurrent weight loss, this novel automated home-cage task improves comprehensive pre-clinical assessment of cognitive functions in mouse models of psychiatric and neurological disorders and enables analysis during short developmental stages. Additionally, running this task in this PhenoTyper home-cage provides the opportunity for combining analysis of discrimination- and reversal learning performance with measures of spontaneous behaviors [2], avoidance learning [4] and anxiety [5].

References

1. Remmelink E, Smit AB, Verhage M, Loos M. Measuring discrimination- and reversal learning in mouse models within 4 days and without prior food deprivation. (Submitted) 2016.
2. Loos M, Koopmans B, Aarts E, Maroteaux G, van der Sluis S, Verhage M, et al. Sheltering Behavior and Locomotor Activity in 11 Genetically Diverse Common Inbred Mouse Strains Using Home-Cage Monitoring. PLoS One 2014;9:e108563.
3. Dias R, Robbins TW, Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. Nature 1996;380:69–72.
4. Maroteaux G, Loos M, van der Sluis S, Koopmans B, Aarts E, van Gassen K, et al. High-throughput phenotyping of avoidance learning in mice discriminates different genotypes and identifies a novel gene. Genes Brain Behav 2012;11:772–84.
5. Aarts E, Maroteaux G, Loos M, Koopmans B, Kovačević J, Smit AB, et al. The light spot test: measuring anxiety in mice in an automated home-cage environment. Behav Brain Res 2015;294:123–30.

Tracking of Individual Mice in a Social Setting Using Video Tracking Combined with RFID tags

James Douglas Armstrong[1,2], Abraham Acevedo Arozena[3], Rasneer Sonia Bains[3], Heather Cater[3]
Agis Chartsias[2], Patrick Nolan[3], Duncan Sneddon[3], Rowland Sillito[2] and Sara Wells[3]

[1]Actual Analytics Ltd, United Kingdom
[2]University of Edinburgh, United Kingdom
[3]Mary Lyon Centre, MRC Harwell, UK.
jdarmstrong@actualanaytics.com

Background

Rodents used in laboratory research are housed in small groups in cages where they eat, sleep, drink, groom and interact socially. Procedures and behavioural tests to analyse an animal's capabilities and fitness are often laborious, slow, subjective and unnatural. Experimenter influence is a particularly difficult issue; even if the data capture itself can be automated, or controlled, the presence of the scientist during the experiment may have an influence [e.g. 1]. These challenges are not new but with increasing interest in longitudinal studies, for example the analysis of effects of aging or the impact of neurodegenerative diseases the ability to accurately and consistently measure behaviour over prolonged times becomes extremely important [2].

We designed, implemented and validated a system for collecting longitudinal data on individual animals who are, importantly, still housed within a normal social group. A range of homecage analysis systems already exists but none quite meet the constraints we had: Most of the existing systems are focussed on single animals or use essentially bespoke environments. Instead we sought to develop a system that was completely compatible with modern high-density IVC caging systems and that could slot seamlessly into high-throughput facilities [3].

System Design

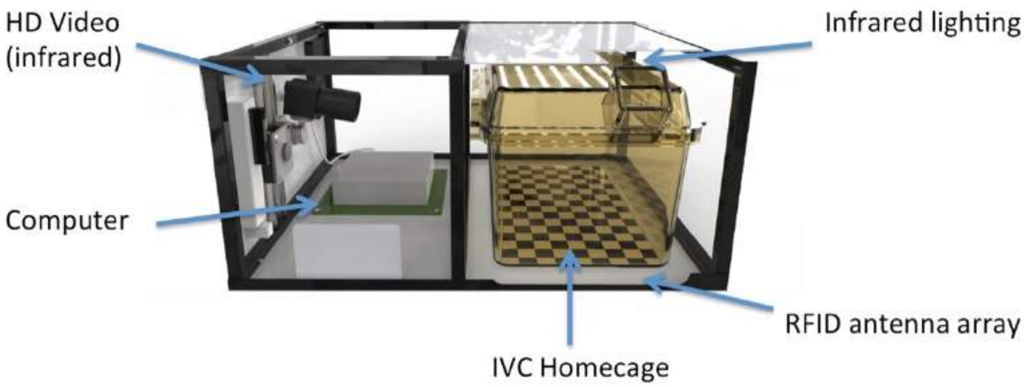


Figure 1: Schematic of the Home Cage Analysis System.

The system is entirely built around a normal IVC homecage designed for a small social group of mice (Fig 2). All the studies described here were performed using Techniplast Sealsafe IVC Blue line cages. The general design is compatible with other cages and racks of similar dimensions and we have also developed a similar system for rats [4]. Identity tags are already widely used in the field and involve the non-surgical implantation of minute, low-cost RFID asset tags into each animal. To achieve spatial monitoring of individual location and detect animal activity, the home-cage is then placed on a low profile base-plate that contains a 2D array of RFID antennae. Each of the antennae in the baseplate is designed to energise a small spatial area within the cage and read the identity of a tagged animal within that space. We also included an infrared light source and infrared

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

camera to record video footage for validation and to allow automated behaviour recognition. A small computer is included to record the data; and a frame to match the rack it is installed into; and the appropriate power supply units that complete the system. The complete physical system occupies two spaces in a standard IVC mouse rack and holds one standard, unmodified cage (i.e. 50% occupancy in a full rack).

Identification of locomotor behaviours of individuals

Locomotion of individuals can be extracted from the social group using the unique ID from the RFID tags. Each antenna under the homecage returns an approximate position and time. To validate the system we recorded a series of top-down videos and manually annotated the centre of mass of each animal once every 25 frames (once per second). We then looked at the correlation of distance travelled from the manual annotations compared to the same distance inferred from the RFID array of the baseplate (Fig. 2). Since the antennae position the animals at the spatial centre of the electromagnetic fields then the distance moved should, on average, be a slight underestimate of actual distance moved but nonetheless shows a strong rank correlation (Spearman's rank coefficient $\rho = 0.911$, $p = 7.55 \times 10^{-16}$, N = 39).

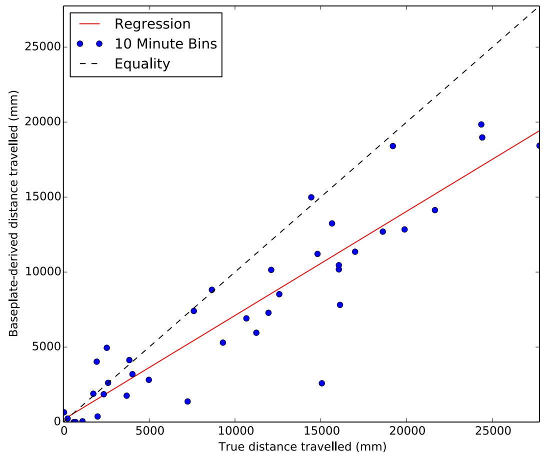


Figure 2. Correlation plot of distance (mm) moved by each animal in 10 minutes measured by annotators using top-down video footage or estimated from the RFID antenna array under the homecage.

When applied to individuals, cages and entire cohorts over multiple days, patterns of activity can be measured accurately. For more details, see the accompanying presentation by Nolan.

Identification of individual behaviours from video

Infrared video provides a means to extract more complex behaviours of the animals at any point in the light-dark cycle. We obtained a series of training videos with a range of common, spontaneous behaviours annotated by three experts. One example is 'drinking'. We trained an SVM classifier to identify drinking based utilising histograms of spatiotemporal features similar to optic flow, extracted at the drinking spout. Bouts of drinking can then be assigned to the most likely animal based on RFID-based position estimates. At the time of writing, drinking events are detected with an average frame-by-frame accuracy of 91.2% and a mean accuracy of ~84.2% when normalised to take into account the low probability of observing drinking events. When applied to a set of more than 40 half hour videos, automated estimates of the proportion of time spent drinking correlate well with human estimates although tending to overestimate with current parameters (Fig 3).

Figure 3 Left: Shows the correlation of drinking bouts as detected by the automatic classifier compared to a human expert. Right: Example frame from video.

Behaviour of animals as a social group in the homecage

The tracking of individuals within the group also unlocks a potential to measure spontaneous social interactions within the cage. While our work on this aspect is at a very early stage we can show clear patterns that naturally emerge. We recover a range of parameters including spatial positions and average pairwise distances between individuals onto a spatial map we can easily visualise where in the cage the animals congregate (huddle). Dividing these data into time bins allows us to see the animals huddled together during the light phase and then splitting up to explore their environment more actively (Fig 4) and independently just in advance of the dark phase (anticipation).

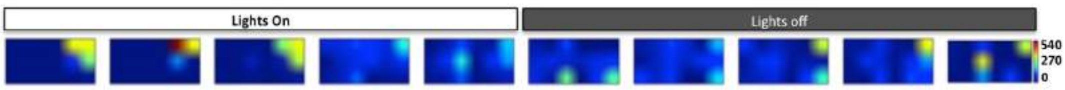


Figure 4. Each frame shows the average position of the animals over a 6minute period coloured by the seconds spent in that position. . The lights are on during the first half (left) and off for the second half (right). The animals were reared under a constant 12:12 light:dark cycle. The social group disperses just before the expected lights off time.

Discussion

The system we present is highly compatible with modern, high-density animal facilities. It has minimal disruption to locally established animal husbandry procedures. The ability to work in normal social groups has profound welfare implications and helps simplify colony management. The combination of HD video and RFID tracking provides the data needed to observer behaviours yet recover the identity of individual animals. Most importantly it allows the automated extraction of a range of spontaneous behaviours (some examples described above) and can be used in longitudinal studies spanning days/weeks. Retaining the data also means that all automated analysis performed by the system can easily be validated. The automatic extraction of behaviours from video is a complex challenge and algorithms are under constant development and refinement. However the platform is stable and the data can be retained so as new algorithms are developed, existing datasets can easily be re-examined. The first studies using these systems are nearing completion and we see very clear patterns of behaviour that vary with the genetic strain under investigation and which will be presented at the workshop.

Mouse Husbandry and Ethical Approval

Animal studies described here were subject to the guidance issued by the Medical Research Council (UK) in Responsibility in the Use of Animals for Medical Research (July 1993), were dependent on an institutional Animal Welfare and Ethical Review Body evaluation and were carried out under UK Home Office Project Licenses #30/2995 and #30/3206.

Availability

Data used in this study including video and annotations will be made publically available under an open license at the time of full publication. Early access may be requested by contacting the authors. The system (hardware and software) is marketed by Actual Analytics Ltd (Edinburgh, UK) www.actualanalytics.com

Acknowledgements

This work was supported by the National Council for the 3Rs (UK) under the Crack-It Scheme.

References

1. Sorge R.E et al., (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents *Nature Methods* **11,** 629–632.
2. Richardson C (2015). The power of automated behavioural homecage technologies in characterizing disease progression in laboratory mice: A review. *Applied Animal Behaviour Science* 163, 19-27.
3. Gates H, Mallon AM and Brown S.D.M. (2011) High-throughput mouse phenotyping. *Methods* **53**, 394–404
4. Redfern W.S., et al (2014) Rodent big brother: Development and validation of a home cage automated behavioural monitoring system for use in repeat-dose toxicity studies in rats. *Toxicology letters* Sept 2014.

Role of Computational Models in Automatic Measurement of Behaviors

M. Pavel and H. Jimison

College of Computer & Information Science
& Bouve College of Health Sciences
Northeastern University, Boston, USA, m.pavel@neu.edu

Introduction

Recent advances in technology facilitated the area of behavioral informatics comprising continuous ambulatory sensing, communication and computation is enabling behavioral scientists' research in monitoring, inferences that can be used to optimize Precision Interventions helping individuals to improve their health behaviors [1]. At the heart of behavioral informatics [2] is the ability to monitor and assess behaviors unobtrusively and continuously so that optimal precision intervention – tailored to individuals and context aware – could be delivered via just-intime adaptive intervention (JITAI) [3]. In a similar manner, behavior optimization approaches require accurate measurement, assessment and prediction of behaviors that are accurate and robust enough to optimize interventions. A key component of the control-system theory approach, shown in Figure 1, is the accurate measurement of behaviors that require computational modeling of all the components, from sensors to the high level of behaviors, including various activities and cognitive decisions. The underlying aspect of this approach is the ability to measure and infer the key variables necessary for closing the loop using computational models.

Computational Models in Measurement

The fundamental notion of measurement is a transformation from physical phenomenon, such as movement or heart beat interval, to a numerical or symbolic representation such that aspects of the phenomenon, for example a change in caloric intake or speed of walking, map into numerical relationships or operations such as addition or multiplication [4, 5]. To ensure that these operations are meaningful requires models of the measurement processes.

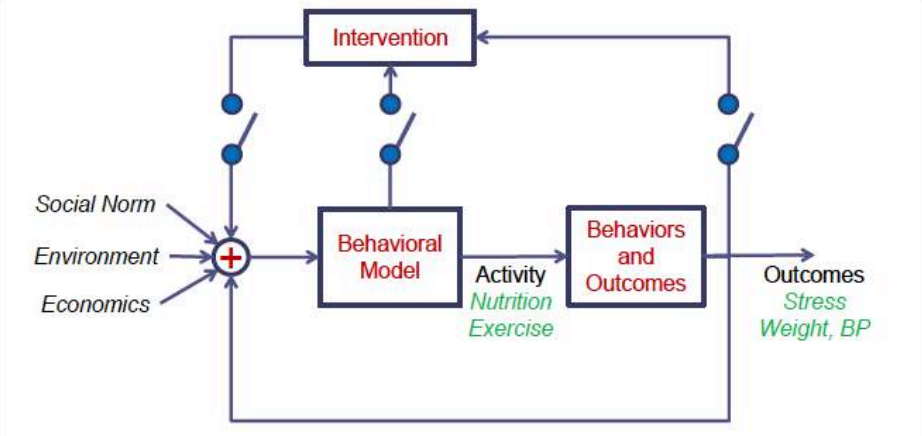


Figure 1. A schematic of the control-theoretic approach to behavioral modification requiring model-based accurate measurement of behaviors

One way to represent the measurement process is as a transformation H from the behavioral variable of interest $\Psi \in \Psi$ to the measured quantity y, i.e., $y(t) = H[\Psi, \mathrm{C}](t)$, where C represents the context as well as random

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

effects such as noise. The measurement transformation is involved in any measurement including physics and biology, but specifying the transformation (measurement model) H is particularly critical in measuring behaviors. The reason is that the observable quantities are typically remotely related to the aspects of behaviors in question – especially in automatic measurement of behaviors. For example, in measuring the intensity (energy expenditure) of activities such as walking, we use accelerometers attached to different parts of the body. But these devices provide only a noisy representation of acceleration in three dimensions that depend on the way the device is attached to the monitored individual. A fairly sophisticated model-based computation is needed to convert these 3D acceleration data to the number of steps in a way that would be independent of the body location of the device [6, 7]. Different transformations are required to convert accelerometer data to sleep duration and sleep quality. Similar considerations must be incorporated for unobtrusive measurement of the speed of walking using stationary sensors in the home of older adults [8]. This example illustrates that the measurement transformation needs to incorporate all the significant noise sources in order to assure unbiased measurements.

A very different example of automatic measurement of behaviors and their interpretations involves the ability to monitor automatically, unobtrusively, and frequently individuals' cognitive activity and fine visually-guided movements [9, 10]. Using data from specifically designed computer games and from individuals' computer mouse movements, it is possible to derive the type of data that reflect individuals' scores on standard neuropsychological tests. Varying various aspects of the games, e.g., number of places to search, and the resulting performance changes can be used to separate individuals' cognitive abilities. In a similar manner, examination of step-by-step performance on a "memory concentration game" enabled us to estimate the momentary working memory capacity of individual players. This inference is made possible by developing and deploying mechanistic, computational models that relate the observed mouse movements to the underlying cognitive and motor processes.

Acknowledgements

Research supported by the National Science Foundation (Grant 5RC1AG036121), the National Institute on Nursing Research (Grant P20-NR015320), the National Institute of Disability, Independent Living, and Rehabilitation Research (Grant 90RE5023), and the National Institute on Aging (Grant P30-AG048785)

References

1. N. Saranummi, D. Spruijt-Metz, S. S. Intille, I. Korhonen, W. J. Nilsen, and M. Pavel, "Moving the science of behavioral change into the 21st Century: Novel solutions to prevent disease and promote health," *IEEE Pulse,* vol. 4, pp. 22-24, Nov-Dec 2013.
2. M. Pavel, H. B. Jimison, I. Korhonen, C. M. Gordon, and N. Saranummi, "Behavioral Informatics and Computational Modeling in Support of Proactive Health Management and Care," *Biomedical Engineering, IEEE Transactions on,* vol. 62, pp. 2763-2775, 2015.
3. I. Nahum-Shani, S. N. Smith, A. Tewari, K. Witkiewitz, L. M. Collins, B. Spring*, et al.*, "Just in time adaptive interventions (JITAIs): An organizing framework for ongoing health behavior support," in *Methodology Center technical report No 14-126*, P. State, Ed., ed. University Park, PA, 2014.
4. D. Krantz, D. Luce, P. Suppes, and A. Tversky, "Foundations of measurement, Vol. I: Additive and polynomial representations," 1971.
5. F. S. Roberts, *Measurement theory with Applications to Decision Making, Utility, and the Social Sciences*. New York: Cambridge University Press, 1985.
6. A. Khan, Y. Lee, S. Lee, and T. Kim, "Accelerometer's position independent physical activity recognition system for long-term activity monitoring in the elderly," *Med Biol Eng Comput,* vol. 48, pp. 1271-9, 2010.
7. J.-S. Sheu, W.-C. Jheng, and C.-H. Hsiao, "Implementation of a Three-Dimensional Pedometer Automatic Accumulating Walking or Jogging Motions in Arbitrary Placement," *International Journal of Antennas and Propagation,* vol. 2014, 2014.
8. S. Hagler, D. Austin, T. L. Hayes, J. Kaye, and M. Pavel, "Unobtrusive and ubiquitous in-home monitoring:

a methodology for continuous assessment of gait velocity in elders," *IEEE Trans Biomed Eng,* vol. 57, pp. 813-20, Apr 2010.

9. S. Hagler, H. B. Jimison, and M. Pavel, "Assessing executive function using a computer game: Computational modeling of cognitive processes," *IEEE Journal of Biomedical and Health Informatics,* vol. 18, pp. 1442-1452, 2014.
10. S. Hagler, H. Jimison, and M. Pavel, "Unobtrusive in-home assessment by means of everyday computer mouse usage," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS,* pp. 3816-3819, 2011.

Automatic Measurement of Behaviors for Intervention Optimization

H.B. Jimison and M. Pavel

College of Computer & Information Science and Bouve College of Health Sciences, Northeastern University, Boston, MA, USA. h.jimison@neu.edu

Introduction

Many countries are experiencing unprecedented demographic changes with an aging of the population. Older adults often struggle to maintain quality-of-life and independence in the presence of chronic diseases and prevailing illness; for low income elders with limited access to a resource-rich supportive environment, these challenges are even more profound. With the dramatic advances in sensors for home monitoring, wearable devices and ubiquitous communications technologies, we have an unprecedented opportunity for delivering effective, tailored, and sustainable health interventions to the home for older adults on a large scale in a cost-effective manner. Figure 1 shows the information flow for a general purpose system for devlivering a health intervention to the home [1]. The various types of sensors that can be used for this type of intervention are illustrated on the right of the diagram. They range from wearable accelerometers, wall-mounted motion sensors, ECG leads, bed sensors to monitored interactions with computers or cell phones. The raw data from these sensors must be filtered and combined with other possible inputs to infer behaviors, context and clinical states (e.g., cognitive function or physical function). In the subsequent sections of this paper we will describe how these inferences are used as inputs to our Health Coaching Platform to tailor semi-autommated messages to both the older adult in the home and other members of a care team, including family members, a coach or care manager, or clinician.

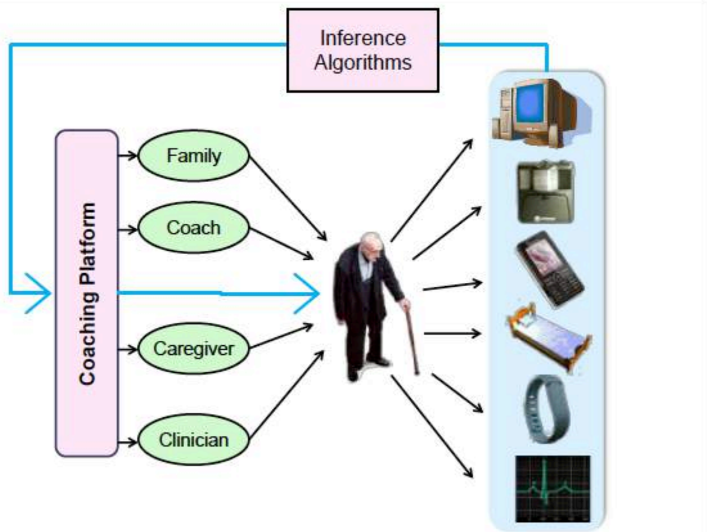


Figure 1. Diagram showing how monitored data from the home can be used to inform a health intervention. Individual sensors collect data from the home and environment. Inference algorithms generate estimates of behaviors, context, and important clinical states. Tailored messages are sent to the patient and care team [1].

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Health Coaching Framework

Most older adults have more than one chronic condition that requires care management in the home (e.g., diabetes, heart failure, etc.)[2]. Health behavior change interventions are an integral part of both disease management, as well as interventions to keep older adults independent, highly functioning, and able to age in place (avoid needing to move to assisted living facilities). Important health behavior change interventions that cut across chronic conditions include physical exercise, sleep management, stress management and socialization. Figure 2 describes our approach to delivering tailored just-in-time health interventions based on monitoring data from the home. Thus far we have developed and evaluated intervention modules for cognitive health[3-5], sleep management[6], socialization[7] and interactive video exercise[8, 9]. All of our intervention modules model known health behavior change variables, such as self-efficacy, motivations and barriers to change, and readiness to change (as shown in the upper left box of Figure 2). The initial patient states from an intake assessment are fed to our Dynamic User Model (center top box of Figure 2). The estimates of patient activities and clinical states derived from the monitored sensor data are also used as continuous input into the Dynamic User Model. With algorithms based on active methods[10] we use the monitoring data to trigger tailored messages of feedback, encouragement and reminders for the older adult in the home. These messages may be edited by a coach if needed. Thus, the monitoring measures are used to infer behaviors, which in turn update the Dynamic User Model to trigger tailored messages to the older adult. The advantage of this approach is that the just-in-time monitoring and tailored feedback make it possible to deploy more effective health behavior change interventions.

Individually-tailored interfaces

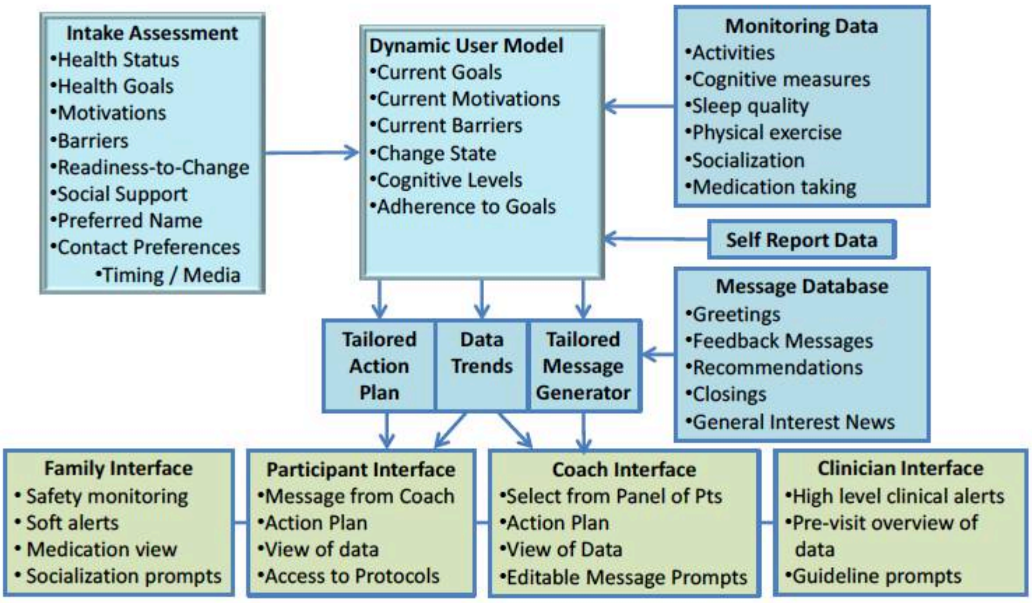


Figure 2. The framework for the Health Coaching Platform includes a dynamic user model based on initial assessments and updated with sensor data. This user model drives the semi-automated tailored messaging for the health interventions, as displayed in tailored interventions, as displayed in tailored interfaces for each stakeholder.

Figure 2 also shows how the same framework can be used to generate tailored interfaces for the older adult and the care team, including informal caregivers, such as family members or home health workers. The patient interface includes summary reports of adherence to their action plan for the week and both graphics and summaries of the monitored data and its implications (e.g., sleep quality).

We have also found that it is important to assess data sharing preferences and explicitly model how best to display data for the family and clinician interfaces in a manner that ensures patient privacy. However, in our studies we have found that older adults are far more willing to share data than their adult children would expect [10]. Data sharing preferences are based on a trade-off between a potential loss of privacy and the potential benefit of being able to age in place and remain independent.

Conclusion

As we move toward an era of changing demographics and a need for cost-effective methods of caring for older adults in the home, the new monitoring technologies, mobile communications devices and algorithms for inferring health behaviors in real time can play an important role in transforming health care. Our technique for using a Dynamic User Model of an older adult within a modular Health Coaching Platform provides a framework for providing tailored just-in-time feedback to older adults and caregivers to improve health behavior change interventions.

Acknowlegements

The authors would like to acknowledge funding support from the U.S. National Science Foundation (Grant 1407928), the U.S. National Institute of Nursing Research (Grant P20-NR015320), the U.S. National Institute of Disability, Independent Living, and Rehabilitation Research (Grant 90RE5023), and the U.S. National Institute on Aging (Grant P30-AG048785).

References

1. M. Pavel, H. B. Jimison, H. D. Wactlar, T. L. Hayes, W. Barkis, J. Skapik*, et al.*, "The role of technology and engineering models in transforming healthcare," *IEEE Rev Biomed Eng,* vol. 6, pp. 156-77, 2013.
2. (February 5). Chronic *Disease Management*. Available: https://www.ncoa.org/healthy-aging/chronic-disease/
3. S. Hagler, H. B. Jimison, and M. Pavel, "Assessing executive function using a computer game: Computational modeling of cognitive processes," *IEEE Journal of Biomedical and Health Informatics,* vol. 18, pp. 1442-1452, 2014.
4. H. Jimison, M. Pavel, and T. Le, "Home-based cognitive monitoring using embedded measures of verbal fluency in a computer word game," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2008, pp. 3312-5, 2008.
5. H. Jimison and M. Pavel, "Embedded assessment algorithms within home-based cognitive computer game exercises for elders," Conference *proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference,* pp. 6101-6104, 2006.
6. J. M. Perez-Macias, H. Jimison, I. Korhonen, and M. Pavel, "Comparative assessment of sleep quality estimates using home monitoring technology," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2014, pp. 4979-82, Aug 2014.
7. H. B. Jimison, K. A. Klein, and J. L. Marcoe, "A socialization intervention in remote health coaching for older adults in the home," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 7025-7028.
8. F. Ofli, G. Kurillo, S. Obdrzalek, R. Bajcsy, H. Jimison, and M. Pavel, "Design and Evaluation of an Interactive Exercise Coaching System for Older Adults: Lessons Learned," *Biomedical and Health Informatics, IEEE Journal of,* vol. PP, pp. 1-1, 2015.
9. S. Hagler, H. B. Jimison, R. Bajcsy, and M. Pavel, "Quantification of human movement for assessment in automated exercise coaching," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 5836-5839.
10. H. B. Jimison and M. Pavel, "Integrating Computer-Based Health Coaching into Elder Home Care," in *Tchnology and Aging*, A. Mihailidis, Boger, J., Kautz, H., and Normie, L., Ed., ed Amsterdam, The Netherlands: IOS Press, 2008.

What Can Watching Old Mice Tell Us About Aging?

S.J. Bonasera[1], N.W. DeKorver[1], T.R. Chaudoin[1], E.H. Goulding[2], A.K. Schenk[3], B. Zylstra[3], C. Hanton-Rye[1] and Y.-J. Kwon[3]

[1]Department of Internal Medicine/Division of Geriatrics, University of Nebraska Medical Center, Omaha, Nebraska, United States of America, corresponding author: sbonasera@unmc.edu.
[2]Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States of America.
[3]Department of Physics, Randolph College. Lynchburg, Virginia, United States of America.

Older adults constitute the most rapidly expanding demographic group across the globe.It is thus imperative to find interventions that improve individual healthspace, lower chronic disease risks, and ultimately improve functional status.

My laboratory has demonstrated that older mice show many of the behavioral hallmarks of aging expected to occur in a clinical population. These age-associated changes include altered energy balance regulation, altered rest-activity state regulation, and impaired mobility. These age-associated behavioral changes occur in a strain-specific fashion, and thus are useful in exploring the mechanism of specific functional declines; for example, energy balance dysregulation or mobility loss.

For example, we note that aged (21-24 mo old) BALB mice demonstrate weight loss and decreased food intake compared to young (2-3 mo old) BALB mice, or aged C57BL/6 mice[1]. This weight loss is about 15% of total body weight, concordant to what is observed in older human populations experiencing unintended weight loss. Much of this weight loss can be traced to changes in the properties of feeding bouts in aged mice; specifically, aged BALB mice demonstrate significantly fewer feeding bouts throughout the circadian day. Of note, there are no appreciable differences in either mobility or locomotion measures between young and aged BALB mice.

By contrast, aged C57BL/6 mice demonstrate significant decreases in mobility and locomotion compared to young C57BL/6 mice, or aged BALB mice. Aged C57BL/6 mice have almost 60% less daily locomotion. Locomotor trajectories of aged C57BL/6 mice also demonstrate more ataxia compared with young C57BL/6 and aged BALB mice. These phenotypes are concordant with what may be observed in older human populations experiencing mobility losses. We also see that aged C57BL/6 mice demonstrate significantly fewer locomotion bouts throughout the circadian day. Of note, there are no appreciable differences in feeding properties between young and aged C57BL/6 mice.

Analysis of gene expression (by microarray) revealed a striking pattern of age-related changes in the BALB hypothalamus and the C57BL/6 cerebellum; two CNS regions associated with feeding regulation and locomotor function, respectively. Briefly, we observed in aged mice a marked overexpression of genes with purported immune and cell adhesion function in both of these CNS regions (compared to young controls). Many of these genes are expressed early in CNS development, and then silenced throughout adulthood.

Using behavioral analysis techniques (including long-term observation in a custom-designed home cage monitoring system), we can begin examining how genes whose regional CNS expression changes with age contribute to specific functional deficits. For example, we note increased age-related expression of the low affinity F_c receptor Fcgr3 in the BALB hypothalamus. Home cage behavioral analysis of young mice lacking Fcgr3 expression reveals a prominent obesity phenotype characterized by a marked increase in adiposity, hypophagia, and decreased activity. Additionally, we noted increased age-related expression of the complement cascade protein C3 in the C57BL/6 cerebellum. Home cage behavioral analysis of young mice lacking C3 revealed decreased locomotor function in a manner similar to what was observed in aged C57BL/6 mice. Finally, we noted a marked age-associated increase in Tlr2 expression in both the BALB hypothalamus and the C57BL/6 cerebellum. Home cage behavioral analysis of young mice lacking Tlr2 demonstrated increased stability and

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

better consolidation of active and inactive states. These studies suggest that combining sophisticated behavioral phenotyping with precise genetic lesions offer the opportunity to dissect complex behavioral phenotypes arising from clinically relevant conditions, such as aging.

Finally, we demonstrate how better understanding of day-today behaviors in mice directly translates into powerful insights regarding the day-to-day behavior of older adults. For example, we show that off the shelf telecommunications technologies, such as SmartPhones and SmartWatches, can be "repurposed" to measure clinically important behaviors such as gait speed, activity, step count, and lifespace, in ambulatory adult populations. We show that these approaches have significant face validity to measure the above clinical constructs, and propose that larger scale functional monitoring of clinical populations at risk for significant functional decline (*e.g*., persons suffering from dementia) may be excellent candidate groups to test this intervention and determine if we can leverage the power of day-to-day behavioral patterns to identify acute and subacute functional changes.

1 All animal studies were performed according to both institutional and national regulations governing animal use. Complete protocols for mouse use are available from the University of Nebraska Medical Center Institutional Animal Care and Use Committee (IACUC). For all home cage monitoring studies, mice received visual health checks on a daily basis, and were given access to water and chow *ad libitum*.

Detecting Animals' Body Postures Using Depth-Based Tracking Systems

P. Pons[1], J. Jaen[1] and A. Catala[2]

[1]ISSI Research Group, Department of Information Systems and Computation, Polytechnic University of Valencia, Valencia, Spain
[2]LoUISE Group, Computing Systems Department, University of Castilla-La Mancha, Spain
ppons@dsic.upv.es, fjaen@upv.es, alejandro.catala@uclm.es

Introduction

The spreading field of Animal-Computer Interaction (ACI) [6] focuses its efforts in developing suitable technology to improve the wellbeing of animals - and also of humans along the way - in many areas: developing digital games for animals in order to avoid boredom and stress [9,15], providing enrichment and stimulation for captive animals in zoos [2,3], facilitating the task of medical assistance dogs [7,12] or developing new means of communication between humans and animals [1,8].

Within ACI research, there is growing interest in recognizing animals' behaviors by means of technology. This challenging task comprises two main questions to address. On the one hand, understanding how behavior can be modeled from observable postures, movements or actions for a specific animal species. On the other hand, defining suitable technology to recognize these specific behaviors attending to the species and the context in which they are produced. In all cases, behavior recognition could provide countless benefits for animal wellbeing as well as for animal-human communication. Being able to automatically recognize an animal's behavior by means of technological artifacts would allow knowledge databases to be created which could serve to study the commonalities of the species and singularities of specific individuals. Machine learning algorithms could also be applied to this information in order to learn and extract patterns on animal behavior in specific scenarios or during their regular daily activities. In this way, changes in behavioral patterns could be detected, alerting the humans responsible for the animal to examine closely the reasons for such transition. This could allow early detection of illnesses or behavioral problems. In a less demanding scenario, learning body postures and behavioral patterns could assist in the development of tailored systems for animals, such as interactive games which adapt to the preferences of the animal who is playing and react accordingly to create an engaging experience [9,10]. As for communication, assigning some meaning to the gestures/postures of the animal which would be automatically recognized by the system would enable them to communicate remotely with a human being via those gestures. This could be the case of Search and Rescue dogs which have to undertake demanding tasks out of the sight of the human handler. Recognizing the behavior of the dog while it is performing its task allows the handler to assess the animal's welfare at all times and to know if the animal is communicating some findings even when he is out of sight [1]. In addition, humans without prior knowledge on animal behavior or blind people who cannot directly observe the animal could understand what the animal is expressing if a system could tell them about it by analyzing the animals' behavior.

This paper will firstly give an overview on existing literature about activity and posture recognition on animals. Secondly, a novel tracking system for cats capable of detecting basic body postures and other relevant information will be presented. Finally, further research questions and extensions to this system will be outlined.

Related Works

Several works within ACI have addressed the necessity of recognizing and identifying animals' postures and activities in different scenarios. These works mostly rely on wearable devices with attached accelerometers and also gyroscopes. Collar-worn accelerometers have been use to detect a variable range of activities: comercial devices such as Whistle® or FitBark® only perform basic activity level recognition, e.g. resting vs. moving, while the work by Ladha et. al. [4] is able to differentiate between a wider range of behaviors on dogs (14 activities and 2 postures). Collar-worn accelerometers have also been used to recognize head gestures on dogs

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

[14], and accelerometers located on the dog's harness have also allowed to differentiate between several static and dynamic postures [1]. A different approach is used in *Canine Amusement and Training* [16], in which instead of using accelerometers, IR emitters are attached to the dog's harness, and a Wiimote's IR camera is placed on the ceiling to detect the location and posture of the animal by tracking the IR emissions.

However, wearable devices could not be used in all situations. Some animals might find their use uncomfortable, it could limit their spontaneous reactions, and it might be risky to use them with protected species or captive animals. Non-wearable mechanisms would allow more natural interactions while guaranteeing the animals' wellbeing. Several non-wearable tracking systems have been developed, such as Poultry.Internet [5], which uses computer vision techniques to detect the location an orientation of a chicken inside a house backyard, but no postures are identified. Another example is Purrfect Crime [13], a game in which a Microsoft Kinect® sensor is used to detect the location of a cat inside the tracked area. No postures are being detected either. Therefore, these works are lacking very valuable information such as the body posture and the place to where the animal was focusing its attention. Nevertheless, a system such as Purrfect Crime, which captures depth information from the scene, would be a very good starting point in order to exploit depth information to enable animals' body posture recognition.

Tracking Animals Using Depth Information

In order to provide a non-wearable system capable of detecting animal's body postures, we have developed a prototype of a depth-based tracking system for cats using a Microsoft Kinect® v1.0 sensor [11]. This sensor provides both color (see Figure 1a) and depth streams (see Figure 1b). Each depth frame provides, for each pixel, the distance in millimeters from the camera plane to the nearest object in that particular pixel (see Figure 1b). Placing the sensor looking down from the ceiling allowed to track a wide interaction area and get aerial images of the cats' body. We recorded images of cats with the Kinect® sensor using this set-up. In the obtained depth images, the contours of the cats can be clearly observed from above. The information provided by the depth sensor, represented as a gray scale image, allows a human eye to easily differentiate between the different body parts and the overall body posture the animal was adopting. Therefore, the main challenge was to develop a system capable of exploiting the information provided by the depth stream in order to automatically recognize those postures.

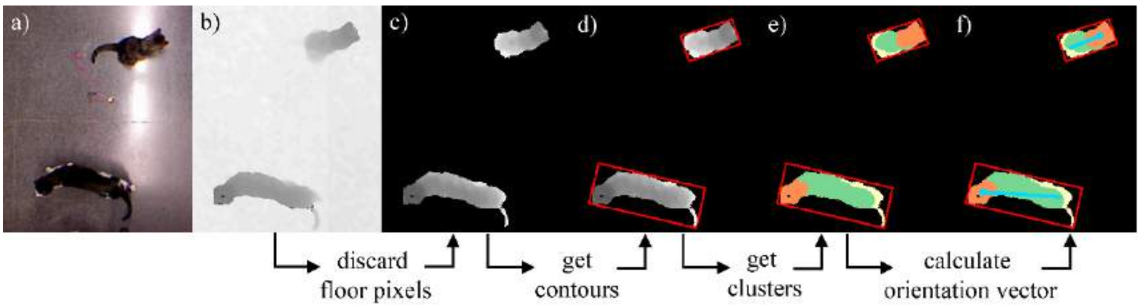


Figure 1. Process of extracting the cat's orientation: (a) color frame (b) depth frame, (c) background segmentation, (d) cat contours, (e) clusters for head, body and tail, (d) orientation vector.

The first processing step of the algorithm consists of extracting the cat's pixels from the depth frame (see Figure 1c). As the depth sensor was placed on a fixed position, our approach to background segmentation simply consisted in removing the pixels whose depth corresponded to the ground. This approach can unintentionally remove the cats' tail (as observed in Figure 1c, cat in the top-right corner of the image), which for some other research could be a source of valuable information when creating richer posture descriptors. Thus, more elaborated approaches, such as combining information from both color and depth images, and even considering a more precise depth sensor, would help to overcome this current limitation. With the floor removed from the image, computer vision algorithms are applied to extract the cats' contours, which now appear as grey-scale blobs in the image (see Figure 1d). In this step, a cat's location within the tracked area can be determined by using the centroid of the extracted contours as a 2D coordinate. Each detected contour is then processed by a clustering algorithm, which groups the pixels by their depth value and relative position (see Figure 1e): pixels of similar depth located together in the image would be grouped within the same cluster. The clustering algorithm divides the cat's pixels into three clusters, which are then classified to determine which part corresponds to the cat's head, body and tail respectively. Having the cluster positions determined, an orientation vector can be defined, from the center of the body/tail cluster to the center of the head cluster (see Figure 1f), to roughly estimate the cat's field of view. Different cat postures can also be differentiated by observing the depth frames and clusters' information (shape descriptors, number of pixels, average depth and location of the clusters). This preliminary version of the tracking system is already capable of classifying the different body parts of each cat as well as detecting several postures, e.g. sitting/semi-sitting, walking/standing, jumping, or turning, using supervised and unsupervised classification algorithms.

Challenges for Depth-Based Behavior Recognition

Animal behavior recognition using depth-based tracking systems seems to be a promising research line. By exploiting the depth information available, it is possible to differentiate between different body parts of an animal within the image attending to their relative location, depth and shape. In this way, an implicit 3D representation of the animals' body can be elicited. This provides a lot more information to differentiate between postures in which analysis of 2D images would not be conclusive. Moreover, these implicit 3D models of postures could be used to construct richer behavior models, ethograms, or activity recognition systems if temporal information and machine learning methods are also used. The ability to recognize the posture, behavior and activity of the animals being tracked could also allow to study how these animals interact with other elements in the environment, such as digital or interactive technology. In this case, ACI studies could benefit from the use of these systems as they could facilitate the study of animals' interaction with technology, helping to improve the design process of animal-centered technology and the animals' wellbeing.

Several issues remain to be studied, such as the reliability and potential of these systems, the behaviors which could and could not be recognized using this approach, and the animal species which these systems could work with. Different animal species might require different set-up configurations and algorithms depending on their skeleton and physiognomy. For example, the tracking approach described in this paper, i.e. blob detection, clustering and classification, could be easily adapted to work with other four-footed animals with similar physical characteristics such as dogs, tigers, horses, etc., provided that the classification algorithm is correctly adapted to fit the new data. A different approach should be taken with other species such as orangutans, in which hand gestures and facial expressions are likely to provide more information. Moreover, birds' behavior seem to be difficult to track using body postures or gestures, however movement and sound patterns could be significant indicators in this case. In addition, the specific relevant behaviors which are going to be detected for an animal species could also determine the position and location of the sensor, e.g. top-down or front-view approach. It can also condition the number of sensors required, for example if the animal moves around an open space area several sensors might be needed. What is evident is that advice from experts in animal behavior will be extremely valuable in the design and development of these systems.

Conclusion and Future Work

This paper presents a tracking system for animals based on depth information which would allow the detection of a cat's location, body posture and orientation. The use of depth information along with traditional computer vision techniques provides more information about the tracked animal than using solely an RGB camera. Hence this system could be a promising starting point towards the automatic detection and analysis of animals' behavior without requiring the animal to use any wearable device. This system is not limited to domestic scenarios, and it can also be used in kennels or zoos, in which the recognition and analysis of animals' behavior could be of great value. The system could be extended to work with other animal species. In addition, more postures and behaviors could be detected by improving and/or combining the classification algorithms and including temporal information from previous frames.

Acknowledgements

This work is funded by the European Development Regional Fund (EDRF-FEDER) and supported by Spanish MINECO with Project TIN2014-60077-R. It also received support from a postdoctoral fellowship within the VALi+d Program of the Conselleria d'Educació, Cultura I Esport (Generalitat Valenciana) awarded to Alejandro Catalá (APOSTD/2013/013). The work of Patricia Pons is supported by a national grant from the Spanish MECD (FPU13/03831). Special thanks to our cat participants, their owners and feline caretakers for their time and dedication. We would also like to thank the reviewers for their insightful comments and feedback.

References

1. Bozkurt, A., Roberts, D.L., Sherman, B.L., Brugarolas, R., Mealin, S., Majikes, J., et al. (2014). Toward Cyber-Enhanced Working Dogs for Search and Rescue. *IEEE Intelligent Systems* **29**(6), 32–39.
2. Carter, M., Webber, S., Sherwen, S. (2015). Naturalism and ACI: Augmenting Zoo Enclosures with Digital Technology. *2nd International Conference on Animal Computer Interaction, Proceedings of the 2015 Workshops on Advances in Computer Entertainment Conference - ACE '15 Workshops*.
3. French, F., Mancini, C., Sharp, H. (2015). Designing Interactive Toys for Elephants. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, 523–528.
4. Ladha, C., Hammerla, N., Hughes, E., Olivier, P., Ploetz, T. (2013). Dog's Life: Wearable Activity Recognition for Dogs. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*, 415.
5. Lee, S.P., Cheok, A.D., James, T.K.S., Debra, G.P.L., Jie, C.W., Chuang, W., et al. (2006). A mobile pet wearable computer and mixed reality system for human–poultry interaction through the internet. *Personal and Ubiquitous Computing* **10**(5), 301–317.
6. Mancini, C. (2011). Animal-computer interaction: a manifesto. *Magazine interactions* **18**(4), 69–73.
7. Mancini, C., Harris, R., Aengenheister, B., Guest, C. (2015). Re-Centering Multispecies Practices: A Canine Interface for Cancer Detection Dogs. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 2673–2682.
8. Mealin, S., Winters, M., Dom, I.X., Marrero-garc, M., Bozkurt, A., Sherman, B.L., et al. (2015). Towards the Non-Visual Monitoring of Canine Physiology in Real-Time by Blind Handlers. *2nd International Conference on Animal Computer Interaction, Proceedings of the 2015 Workshops on Advances in Computer Entertainment Conference - ACE '15 Workshops*.
9. Pons, P., Jaen, J., Catala, A. 2015. Envisioning Future Playful Interactive Environments for Animals. Nijholt A, editor. *More Playful User Interfaces,* Springer, 121–150.
10. Pons, P., Jaen, J., Catala, A. (2014). Animal Ludens: Building Intelligent Playful Environments for Animals. In: *Proceedings of the 2014 Workshops on Advances in Computer Entertainment Conference - ACE '14 Workshops*, 3:1–3:6.
11. Pons, P., Jaen, J., Catala, A. (2015). Developing a depth-based tracking system for interactive playful environments with animals. *2nd International Conference on Animal Computer Interaction, Proceedings of the 2015 Workshops on Advances in Computer Entertainment Conference - ACE '15 Workshops*.
12. Robinson, C., Mancini, C., Linden, J. Van Der, Guest, C., Harris, R. (2014). Canine-Centered Interface Design: Supporting the Work of Diabetes Alert Dogs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 3757–3766.
13. Trindade, R., Sousa, M., Hart, C., Vieira, N., Rodrigues, R., França, J. (2015). Purrfect Crime. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15,* 93–96.
14. Valentin, G., Howard, A., Jackson, M.M. (2015). Towards a Canine-Human Communication System Based on Head Gestures. *2nd International Conference on Animal Computer Interaction, Proceedings of the 2015*

Workshops on Advances in Computer Entertainment Conference - ACE '15 Workshops.
15. Westerlaken, M., Gualeni, S. (2014). Grounded Zoomorphism: An Evaluation Methodology for ACI Design. *Proceedings of the 2014 Workshops on Advances in Computer Entertainment Conference - ACE '14 Workshops*, 5:1–5:6.
16. Wingrave, C.A., Rose, J., Langston, T., LaViola, J.J.J. (2010). Early explorations of CAT: canine amusement and training. *CHI '10 Extended Abstracts on Human Factors in Computing Systems,* 2661–2669.

Animal-Human Digital Interface : Can Animals Collaborate with Artificial Presences?

Naohisa Ohta[1,2], Hiroki Nishino[1], Akihiko Takashima[2] and Adrian David Cheok[1]

[1]Imagineering Institute, Iskandar, Malaysia. nohtaster@gmail.com
[2]Graduate School of Media Design, Keio University, Yokohama, Japan

Abstract

This paper presents research (work-in-progress) on animal-computer-human interactions. We explore how a computer system can actively mediate between animals (dogs and cats) and humans, aiming for enabling animals to collaborate (interact) with an artificial presence system, as well as with humans through the system. A concept of the system consisting of multi-sensory and active agent functions is discussed. We describe research plans to investigate how animals behave interacting with virtual reality systems and robot agents. An interactive video interface is used to investigate how the closed visual feedback loop (with partial depth cue perceptions) affects animals' behaviors. For interaction with robots, we investigate if we can establish emotional attachment between dogs and robots.

Introduction

Animals (dogs and cats) watch TV and sometimes seem interested in the displayed content [1-3]. A variety of visual media have been used as artificial visual stimuli in animal behavioral research. Some of them demonstrated that animals are able to treat images as real stimuli [4,5]. Other studies have shown that we can use life-size video images successfully for examining dogs' social communicative abilities [6-10]. Zeil [11] discussed that depth cues may be important factors in visual artificial stimuli so that animals respond to video images in the similar way as they do to real stimuli.

A robot agent is another interesting artificial presence that affects animals' behaviors. Animal-like or human-like robots will offer new possibilities in animal computer interactions. For example, in experiments testing if dogs recognize robots as social partners, it was found that dogs responded significantly better to dog-like furry robots than to remotely controlled cars [12]. Another research demonstrated that dogs appeared to develop social links with robot agents (showed sociality), which did not resemble a human, after having observed social interaction between the robot and a human [13].

The above previous animal research on the effects of artificial stimuli suggests that artificial presences could be used as part of animal-human digital interfaces in general. However, we do not know what essential factors are really effective and how much they contribute for a particular kind of animals. There seems to be no systematic research challenging this topic.

This paper describes our research (work-in-progress) to explore how artificial presences (focusing on virtual reality and robot) can actively mediate between animals and humans, aiming for enabling animals (focusing on dogs and cats) to collaborate with humans through the digital interface. Our final goal is to develop a system through which animals and humans conduct a collaborative task consisting of multiple interactions with the interface. The collaborative task is defined as work such as rescue or playful experiences such as new type sports. We introduce our concept, methods and some experimental procedures.

Concept

If a digital interface can induce animals to keep up their interest in artificial stimuli and to pro-actively behave against the stimuli, then we could go a step forward towards the goal to induce animals to collaborate with the

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

system and with humans, directly or indirectly through the interface. To design such a system, there are many related functions to be considered including animals' multi-modal cognitions.

One of the issues is that animals only keep up their interest in the artificial presence for a short period of time, even if the stimuli are realistic images of the animal's owner, or even if the stimuli are mouse-sized fake-fur covered toys [14]. So it might be difficult for animals to carry out collaborative tasks which may consist of a series of multiple interactions with an artificial presence. As an example, animals are interested in a Skype communication only for a short period of time even if they initially pay close attention to it. This is presumedly due to the lack of multi-modal presences such as olfactory or tactile sense, and appropriate active interaction schemes for animals [15]. In case of PetChatz [16], which is a commercial product, pets pay attention to and stay interested in the system because it has a remote feeding function. This implies that the system should work actively, not just communicating through multi-sensory information channels, but also through changing artificial stimuli interactively to keep the animal's interest up, and finally to induce animals to interact continuously.

We assume a general system model which would consist of multi-sensory communicative functions and active agent functions to affect animals' social behaviors.

Method

We plan to investigate which factors of artificial presences are effective for catching animals' attentions, to keep up their interest, and to induce them to interact consecutively focusing on characteristics of 1) virtual reality systems and 2) robot agent systems. We describe two research areas below.

- Animal's Behavior and Virtual Reality (VR)

Our long-term research will encompass a wide field of VR systems. We will explore how each sensory impression (visual, auditory, olfactory, tactile, and taste) and their combinations affect animals' cognition and behaviors.

Currently, we are focusing on vision in VR, espcially interested in the effect of the closed visual feedback loop as part of a depth perception. Depth perception is an important factor in vision for animals' behavior but usually limited in 2D video images. As Zeil discussed [11], the lack of depth cues and the closed visual feedback loop are potential limitations of video stimuli.

We use an interactive video interface to investigate how the closed visual feedback loop (with partial depth cue perceptions) affects animals' behaviors expecting that the visual feedback would contribute to get animals' attentions and keep their interest up in 2D images. The interface is inspired by our observation that animals are sometimes watching outside through a glass window. From a conceptual point of view, the interface is a subset of "Fish Tank VR" which provides an interactive stereoscopic 3D workspace in front of the display with headcoupled perspective rendering [17]. However, in our study the system's purpose is different from that of Fish Tank VR. We design the interface to provide visual stimuli to animals with head-coupled perspective changes so that they perceive the displayed images as if they see outside through a glass window.

We plan to compare animals' reactions to a video content displayed on a normal 2D display with those to the same content displayed using the interactive video interface. We actually designed the interface considering the following conditions for animals: 1) to provide a smooth feedback that would work at a higher frame-rate (aiming at 60 frames per second), 2) to allow animals move freely without wearing any uncomfortable devices for visualization or sensing.

For the system implementation, we used a game engine, Unity [18]. We controlled a virtual camera in Unity to change the perspective according to the viewer's head position measured by a depth sensing mechanism such as Kinect [19]. As far as video quality is concerned, we considered factors such as spatial and temporal resolutions given in [20]. In our case, we need more display resolution, as the animal usually gets close to the display during experiments. So we use higher quality displays such as 4K (3840 x 2160 pixels) resolution with a frame rate of 60 Hz. While the animal is in front of the display within a certain region, the system detects its head position (horizontal, vertical and depth information), and then it can change the image perspectives interactively with the animal's movement. This could be achieved without using HMDs (head-mount displays) if we do not stick in a stereoscopic 3D visualization.

For the video content, we shoot an HD and/or 4K video with an appropriate view angle which is calculated by the viewer's position and the display size. We assume that the distance of subjects in the video is relatively further compared with the distance between the viewer and the display. In that case, steropsis in the video is not so important as a depth cue [11]. We decided not to use stereopsis as a depth cue in visualization of the interface. This specification finally helped to keep a high frame rate of images given to animals' eyes.

- Animal's Behavior and Robot Agent

The recent study by Lakatos et al. suggests that dogs understand social 'cues' even if it is given by robots, under certain conditions [13]. Such investigation has to be considered in order to understand emotional and/or social relationships between robots and animals. We consider this aspect as very important, since it is expected that our society will utilize autonomous agents such as robots and drones in the very near future. For instance, Softbank, a Japanese telecom company, already began selling their social robot, Pepper, commercially in 2015 [21]. In 2016, the MIT developed robot for home-use JIBO is planned to be shipped [22] and Amazon expects to utilize drones for delivery by 2017 [23]. An investigation of how animals can and will socially react to such autonomous agents physically integrated in our society can be one of the key factors for both human and animal welfare. It may also contribute to help guide dogs and rescue dogs with collaborating and interacting with such autonomous agents. Currently, we are particularly interested in the establishment of emotional attachment between dogs and robots, and investigating in what conditions dogs are inclined to help robots in danger.

Conclusion

We presented a research plan (work-in-progress) on animal-computer-human interaction aiming at enabling animals to interact with artificial presence systems such as VR systems or robots. For VR, we explore key factors to keep animals' interest up in artificial visual stimuli focusing on the effect of the closed visual feedback loop as part of a depth perception. We also investigate if we can establish emotional attachment between dogs and robots. We are still developing this research and looking for any collaboration with people who are interested in the related topics.

References

1. Hirskyj-Douglas, I., Luo, H., Read, J.C. (2014). Is My Dog Watching TV? NordiCHI 2014
2. Why Do Dogs Watch-and React to-TV? (2015) <http://news.nationalgeographic.com/news/2015/01/150111-animals-dogs-television-pets-science-tv-behavior/ > Accessed 22 January, 2016
3. DogTV (2015) < https://dogtv.com > Accessed 22 January, 2016
4. D'eath, R.B (1998). Can Video Images Imitate Real Stimuli in Animal Behaviour Experiments? Biology Review, 73, 267- 292.
5. Bovet, D., Vauclair, J. (2000) Picture Recognition in Animals and Humans, Behavioural Brain Research 109 , 143-165
6. Pongrácz, P., Miklósi, Á., Doka, A., Csányi, V. (2003). Successful Application of Video-Projected Human Images for Signalling to Dogs. Ethology 109 , 809-821.
7. Miklósi, Á., Pongrácz, P., Lakatos, G., Tópal, J., Csányi, V. (2005) A Comparative Study of the Use of Visual Communicative Signals in Interactions Between Dogs (Canis familiaris) and Humans and Cats (Felis catus) and Humans. Journal of Comparative Psychology, 119, No.2 179-186.

8. Tégrás, E., Gergely, A., Kupán K., Miklósi, Á., Tópal, J. (2012) Dog's Gaze Following Is Tuned to Human Communicative Signals. Current Biology, 22, 209-212.
9. Péter, A., Miklósi, Á., Pongrácz, P. (2013) Domestic Dog's (Canis familiaris) Understanding of Projected Video Images of a Human Demonstrator in an Object-choice Task. Ethology 119, 1-9.
10. Müller, C., Schmitt, K., Barber, A. (2015) Dogs Can Discriminate Emotional Expressions of Human Faces. Current Biology 25 , 601-605.
11. Zeil, J. (2000) Depth cues, behavioural context, and natural illumination: some potential limitations of video playback techniques, acta ethologica, 3, No. 1, 39-48
12. Kubinyi, E., Miklósi, Á., Kaplan, F., Gácsi, M., Topál, J., Csányi, V. (2003) Social behaviour of dogs encountering AIBO, an animal-like robot in a neutral and in a feeding situation, Behavioral Process 65 231-239
13. Lakatos, G.,Janiak, M., Malek L., Muszynski, R., Konok V., Tchon, K., Miklósi, Á. (2014) Sensing in dogs: what may make an interactive robot social? Animal Cognition 17 , 387-397.
14. Hall, S., Bradshaw, J., Robinson I. (2002) Applied Animal Behavior Science 79 , 263-271.
15. Goldbeck, J., Neustaedter, C. (2012) "Pet Video Chat: Monitoring and Interacting with Dogs over Distance", CHI 2012, May 5-10, Austin Texus, 211-219.
16. PetChatz Home Page (2015) < http://www.petchatz.com> Accessed 22 January 2016.
17. Ware C. Arthur K. Booth K.S. (1993) "Fish tank virtual reality", INTERCHI'93 Conference Proceedings, 37-42.
18. Goldstone, W. (2011). Unity 3.x Game Development Essentials, Packt Publishing Ltd.
19. Microsoft Kinect - Microsoft SDK(2016). < http://wiki.etc.cmu.edu/unity3d/index.php/Microsoft_Kinect_-_Microsoft_SDK> Accessed 18, March 2016
20. Oliveira, R.F., et al. (2000) Considerations on the use of video playbacks as visual stimuli: the Lisbon workshop consensus, acta ethologica , 3 , 61-65.
21. Aldebaran. SoftBank to Launch Sales of 'Pepper' (2015) < https://www.aldebaran.com/ja/softbank-tolaunch- sales-of-pepper>Accessed June 18 2015.
22. Jibo (2015) < https://www.jibo.com> Accessed June 18 2015.
23. Eadicicco, L., (2016) "Amazon Reveals New Details About Drone Deliveries", Time (Jan 19, 2016) <http://time.com/4185117/amazon-prime-air-drone-delivery/ > accessed 25th January 2016.

The Ethics of How to Work with Dogs in Animal Computer Interaction

I. Hirskyj-Douglas and J. C. Read

Department of Computing, University of Central Lancashire, Lancashire, England.
Ihirskyj-douglas@uclan.ac.uk

Abstract

When designing technology for dogs, ethics need to be considered. The technological approach with dogs has changed with a progressive viewpoint of where a dog is situated within a computer system and how the researcher reacts and treats the dog. This work presents a list of currently used ethical protocols within Dog Computer Interaction (DCI) composed by the Animal Computer Interaction Design (ACID) group to ensure the best animal welfare and the usage of appropriate data collection methods. This list is extended into a discussion on how these principals were formed through a literature review of ACID studies. This paper suggests that the researcher is situated in the tension between human and dog centered design. The extent to which the researcher views the animals will directly affect the results and methods created. It is through the researchers stating how they conceptualise dogs, and being aware where the dog-human relationship place can be situated that more depth is given to the findings.

Introduction

Animal Computer Interaction (ACI) is a stream of research around animals being involved in technology. This research involves the design of interaction modalities and the building of methodological frameworks. Off-shooting theories encompass these challenges with interpretations and outcomes. Throughout and within the background of these factors, the ethical story of the relationships between humans, machines and non-human animals is being played out. This narration not only concerns the manipulation and usage of applications for the ethics involved, but the implications for the participants and more widely for the growing field of ACI. While primarily ACI research focuses on the building and implementation of systems, the ethics of the researcher's relationships with the other entities often is not implicitly discussed despite the results often emerging through the intra-actions of this relationship [3]. This agential realism stance to ACI is shaped into an animal centered design standpoint relationship between animal and human: but to what benefit? This relationship-dance that we have with our pets [10], and more recently the added partner of machines is not a singular interactive lone process. Through an ethical analysis of these complex connections, interpretations can go deeper.

This paper aims to situate the authors' ethics of working with dogs and machines which informs their animal centered design stance. This stance will then be presented with a discussion around the tension between non-human animal, machines and data gathering. Evidence of this discussion in practice and findings are then reflected upon and evaluated. This is then summarized into a belief statement on how to work with dogs in the ACI field.

Ethics of Working with Dogs and Machines

When designing technology for dogs, consideration has to be taken of the ethical practice of the study(s). This consideration of the application is not only for the dogs' welfare, particularly as they cannot explicitly give consent, but also to ensure as precise results as possible. This statement is particularly relevant when gathering dogs' opinions and emotions which could both be influenced through trained behavior. Notwithstanding, animal rights have set boundaries for ethical protocols. This has directly influenced ACI in disbanding speciesism in an argument that ACI inherently includes Human Computer Interaction (HCI) and providing a bases of non-harm towards the non-human animal involved [13]. The ethics of the environmental influence have been acknowledged in ACI within works which aim to shape playful environments through labelling parameters [14]

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

and questioning the dog-machine relationship bond [20] to see how they influence the results. The most notable work is by Vaataji [21] who looks at ACI through the 'three Rs' of cost benefit approach (replacement, reduction and refinement) followed by a literature review of identified ethical guidelines. Her work however is situated in the design of the study and in the reporting of methods used.

Even though the most common form of ACI is with dogs, Dog-Computer-Interaction (DCI), [8; 11; 16] there has yet to be a set of guidelines about conducting studies specifically with pet dogs. Research in DCI covers a number of different topics, such as dog-monitor communication [12; 8], dog-tablet communication [24; 16], dog-unidentified object (UMO) [20], dog tracking [23], aiding human and/or animal communication [14], welfare, and gaming [15]. While the technology used, and motivations behind this use are diverse, they are all within the same species. Despite this DCI creators will often choose ad-hoc or pre-existing Human Computer Interaction (HCI) methods based upon a number of different principals. These principals range from what is available, what is ethical and where the researcher feels situated in their research. However, this approach does not allow for a community to learn from the gathered knowledge. In order to enrich this field and hopefully start building up a set of universal guidelines in DCI, we present below a list of currently followed guidelines by the Animal Computer Interaction Design (ACID) group at the University of Central Lancashire. The ACID group is a research group looking at designing DCI. These guidelines here are presented to not only benefit the DIC community but also benefit those working with dogs in general HCI.

Whilst these guidelines have been used in previous studies by the ACID group, they have not been implicitly laid out. Researchers, however, such as Vaataji [21], have touched upon these in a generic ACI sense: such as adequate housing (related to guideline 3) and forming consent (guideline 1). These guidelines, nonetheless, have mostly been formed through a merger of the researchers own work within ACI, the researchers own opinions on animal-human-computer relationship, and the current guidelines of working with animals.

Whilst the guidelines below are presented as being specific towards dogs, these can, and should where possible, be applied to other researcher in ACI. For example, providing a safe place variants between animal species and should be accordingly adjusted. In a similar route though, animals which are not able to work away through being immobile should be empowered through an alternative route to still enable a form of consent. In this sense the guidelines should provide just that: a guide. Each individual study will need its own set of guidelines but by looking at the customary instances below, this could help open up questions and responses

Similarly to the application of these guidelines to multi-species, guidelines specifically for children (Child Computer Interaction (CCI)) have been developed in HCI [5]. These can often be mapped to animals due to the similar challenges of children and animals face with technology of being none or limited verbally especially with abstract concepts and actions [6]. An instance of this is by giving children better control and trust [9] through familiarising them with the lab, which guideline 1 & 2 (below) aims to circumvent with animals.

ACID Group Guidelines for DCI

1. Enable Consent: Walking away. The dog used in any study must be able to walk away from the study as a low-level form of consent and freedom. This not only empowers the dog but also gives more insight into the technology as the dog is not forced to use it.

2. Providing a Safe Place. This can take the form of a dog bed or a blanket which the dog is familiar with if working outside of the home. This is to enable the dog to not only walk away but to have somewhere they feel comfortable to go to. This helps to minimalize any stress felt upon the dog.

3. Work Where Possible Within the Dog's Own Home. This method allows for dogs to both feel comfortable and display normal behaviors as they are familiar with the setting.

4. Have the Owner or Carer Observing. While there is no doubting the benefit of a trained professional in dog-behavior to spot any potential long or short term emotional trauma, the holder of the dog often knows habitual behaviors and will be able to quickly spot and identify any adverse effects. This also has the benefit of enabling the dog to seek comfort if it feels the need to. This enables the dog to be more within the center of the design process through helping the dog's behavioral reactions to shape the research process.

5. When Using Audio or Video Never Show Familiar or Distressing Footage or Sound. In DCI researchers should avoid creating confusion and distress to the dog by showing them audio/visual representations of people or non-human animals they know. This is to prevent disembodiment causing confusion from the lack of understanding around technology on the dogs' behalf. Within this area, the researcher should also not show anything which can cause distress to the dog as this goes against their welfare.

6. Dogs Should Have No Emotional Behavioral Problems. Unless the aim of the study is to improve the emotional behavioral response, generally no dog with behavioral problems along the emotional spectrum should be included. The reason for this is twofold: firstly, the chance of a possible negative effect upon the dog is greater due to an instability of behavior and secondly, there would be some distortion of the results with abnormal behavioral responses.

7. Dogs Should Not Be Trained To Use a System. When working with technology systems it is possible to train the dog to use the system. This however gets the dog to use a system in a trained method and does not collect ordinary results. By allowing the dog to use the system naturalistically the dog will present normal behavior. This, in turn, enables commonplace data collection. The methods and theories then drawn from the gathered data can then be more focused around the center of the dogs' true needs as the results are truer to the dogs ordinary behaviour. This guideline however is not always suitable towards creating systems within working dogs who are relied upon to behave in a certain manner as part of their work.

While these are general principles that ACID researchers follow, and these ideals have been spoken about in various previous studies [11; 21] and ACI conference topics (Bad design ACI@BHCI), the implications behind these decisions have yet to be debated. Even with these things in place does the dog ever really have a choice behind the activities it participates in? Do these principals help a dog choose what it does? Dogs have been shown to follow human choices, gaze [20] and recognize emotions [2], and befriend UMOs [1] and aid each other [18]. This is bringing forward increasing evidence that the interconnected relationships held within and throughout the technology and the environment, influence the study findings. This is why it is so important to record more base statistics of an experimental set-up, as suggest by Vaataja [21], to fully scope the relationships' influences not only on animal-nonhuman animal behavior but also the other agencies involved. Following this, by making the dog as comfortable as possible it both places value on its welfare whilst supporting higher quality research. The ethics of a researcher is always an integral part to the data that is collected with their views and themselves forming part of the research findings [3].

Animal, Machine and Human Relationship

The social attitude behind what a dog is, has evolved drastically, impacting upon the ethics of studies done with dogs. This attitude change has been formed from the knowledge we have gained in relation to the cognitive abilities of dogs, from aiding us in sports and entertainment, to providing companionship, security and physical wellbeing [7]. As a consequence, the abilities of dogs have been demonstrated resulting in dogs being reclassified in 2014/15 as sentient beings rather than as property, having the same legal protection as children under the law in France, Quebec and New Zeeland. This change has been backed up by comparative neuroscience which has found that many of the same things that activate the human caudate also activate the dog's caudate leading to functional homology [4]. While biologists such as Darwin long suspected that all animals have emotions on some scale, this view has only recently been taken seriously as advancements in neurobiology and chemistry have added data. The notion that dogs can experience emotions, like love and attachment, puts dogs on a similar sentience to that of children. This directly impacts on how DCI researchers work with dogs. This reclassification has been acknowledged in some ACI work, such as in the creation of a Doggy Ladder of Participation to design with the dog rather than for the dog; thus appreciating the dog's cognitive aptitude [11].

This reclassification highlights the issues facing researchers working with animals: the problems between gathering data and the ethical morality of working with animals. This is especially true in ACI where for the best part the studies are for animal welfare. If there were no ethics behind working with animal researchers could undoubtedly collect more data, but the accuracy of this data would not be as dependable due to interfering factors (tiredness, motivation etc.). Unlike typical animal experimentation within DCI there should not be a typical cost benefit of harm vs. benefit analysis done as it is not morally acceptable to harm a sentinel being. This is the atypical view HCI researchers hold for children. This is because even with the above mentioned precautions in place a dog can never fully consent to taking part within a study. Due to this, responsibility of welfare between the person and the dog lies with the researcher's country's legislation and more importantly the researcher's point of view on animal welfare and moralities. This no-harm approach, however, is idealistic as not all risks can be foreseen. In order to advance ACI/DCI studies have to be done that can potentially put the animal at risk but will in the long term aid animal welfare. It is, therefore, up to the researcher to find a mid-point which reduces the possibility of risk, insuring the correct welfare of the animal whilst still having usable data. This can be done by placing the animal in the center of the research through a triangulation of knowledge and involvement from the animal's owner(s), animal behaviourist(s) and the researcher(s). This ensures that the animal's welfare is observed from varying perspectives guaranteeing the most amount of gathered knowledge.

Study Reflection

Giving further insight into this, below is a summary of the ethical approach taken within the ACID studies, and the findings that helped informed future ethics from these studies.

Dog Species Appropriate Visual/Audio Stimuli. [10]
This initially investigated the content a dog pays attention to visually and formed the foundation of ACID ethics. Here it was noted that the dog would follow the visual cues of the human owner and thus the owner was instructed not to watch the visual stimuli. Data here was also collected within the dog's own home due to more normalised behavior being displayed by the dog vs in a lab where the dog was more prone to show exploratory behavior (guideline 3). Due to this, the dogs involved often walked away from the media, but this only showed disinterest within the stimuli so was not seen as null data. Within this study, the owner was also involved to ensure the welfare of the dog(s), especially when diverse forms of media were used which could possibly have had a negative effect (guideline 4). It was decided early on that dogs should not be shown media that would be distressing, which it was noted happened with media (sound/images) of known animals (guideline 5). This was further avoided by making sure that no dog involved within the study had behavioural problems to ensure that both the results were normalised and the animals own welfare (guideline 6).

Dog Head Tracker [9]
This study created a head tracker to monitor within three sections (left/right/center) where a dog was looking visually on a screen. The key ethical issues within were twofold: firstly, whether or not to train the dog to sit still to allow for eye tracking and secondly, failing this, to attach a device to the dog's head. There is tension in eye tracking as the most accurate results are at the pixel level which requires more in-depth methods (eye to gaze to face to body) which puts more restraints on the dog thus affecting normal behavior (guideline 7). Whilst Somppi [19] chose to train the dog to look at the screen to collect pixel-level tracking, the dog was not allowed to look away from the screen (trained to position head on a headrest) causing data collection even when the dog was not visually interested. Alternatively, researchers in Lincoln [22] took a different approach that allowed the dog to have freedom over movement but strapped a device onto their face. Even though positive reinforcement was used to get the dog used to wearing the apparatus, this training would have had an impact upon the visual patterns of the dog. This highlighted the tension between accuracy vs. normality and resulted in ACID researchers deciding for a contact-free approach to allow for normal behavior but having the effect of reducing data collection. This approach has since been duplicated in ACI with cats [17] relying on a computer system to analyse the position of the face/body to give approximation. As this study was the first to take place within a lab (due to the technological requirement) it was at this point that the need for a safe place was highlighted to enable

the dog to retreat away from the study (watching media) (guideline 1 & 2). This did result in null data being collected (dog going to bed) but this ensured the dogs welfare.

Concluding Statement

When designing technology for dogs there are numerous factors to consider and with an evolving landscape of what it means to be a dog, the ethics of how to work with dogs plays a vital role in DCI. The undertone of DCI ethical considerations, is the jurisdiction that people aim to empower the dog through and with technology. This animal centered design approach, however, has to initially start with the researcher and is based around their opinions of what a dog is to both them and their research. This placement upon the scaler line of people vs. dogs interaction world view homogony goes beyond the interaction into the results shaping the data and shaping the found knowledge. In order to situate research, the authors here present ACI researchers with a challenge to state within the research conducted what their position is between the animal-researcher and how the researcher sees the animal. This would shape the parameters around the relationship and thus the intra-action taking place. Lastly, DCI guidelines have been coined to ensure the empowerment of ethical working with dogs within ACI. This will hopefully be built upon to become standard practice within DCI.

References

1. Abdai, J., Gergely, A., Petro, E., Topal, J. and Miklosi, A. (2015) An investigation on social representations: Inanimate agent can mislead dogs (canis familiaris) in a food choice task. PLoS One 10(9).
2. Albuquerque, N., Guo, K., Wilkinson, A., Savalli, C., Otta, E., & Mills, D. (2016). Dogs recognize dog and human emotions. Biology Letters, 12(1), 20150883.
3. Barad, K. (2007) Meeting the universe halfway: quantum physics and the entanglement of matter and meaning. Durham, North Carolina: Duke University Press.
4. Berns, G (2013). Dogs are people, too. New York Times.
5. Bruckman, A., Bandlow, A., & Forte, A. (2002). HCI for kids.
6. Druin, A. (2002). The role of children in the design of new technology.Behaviour and information technology, 21(1), 1-25.
7. Godfrey, A. (2006). Human-animal interaction - the place of the companion animal in society. Veterinary Nursing. Butterworth Heinemann Elsevier.
8. Golbeck, J., Neustaedter, C. (2012) Pet Video Chat: Monitoring and interacting with dogs over distance. Proc. CHI EA '12. ACM. 211-220.
9. Hanna, L., Risden, K., & Alexander, K. (1997). Guidelines for usability testing with children. Interactions, 4(5), 9–14
10. Haraway, D. (2003) The Companion Species Manifesto: Dogs, People, and Significant Otherness. Chicago, Prickly Paradigm Press.
11. Hirskyj-Douglas, I., Read, J.C. and Cassidey, B. (2015) Doggy Ladders of Participation. BHCI'15. Animal Computer Interaction.
12. Hirskyj-Douglas, I., Huailan, L., Read,J.C. (2014). Is my dog watching TV? NordiCHI'14. Animal Computer Interaction Workshop.
13. Hirskyj-Douglas, I & Read, J.C. (2013) Animal Computer Interaction Design. ACM-W. Manchester University.
14. Lemasson, G., Duhaut, D & Pesty,S. (2015) Dog: Can you feel it?. BHCI'15. Animal Computer Interaction Workshop.
15. Mancini, C., Lawson, S., van der Linden, J., Häkkilä, J., Noz, F. and Juhlin, O. (2012) Animal-computer interaction SIG. In: ACM SIGCHI Conference on Human Factors in Computing Systems , 5-10 May 2012, Austin, Texas.
16. Paldanius, M., Kärkkäinen, T., Väänänen-VainioMattila, K., Juhlin, O., Häkkilä, J. (2011) Communication technology for human-dog interaction: exploration of dog owners' experiences and expectations. In Proc. CHI'11, ACM, 2641-2650.
17. Pons, P., Javier, J. and Catala, A. (2015) Developing a depth-based tracking systems for interactive playful environments with animals. 12th International Conf. on Advances in Computer Entertainment Tech. 2nd International Congress on Animal Human Computer Interaction.
18. Quervel, Chaumette, M., Dale,R., Marshall-Pescini,S. and Range, F. (2015). Familiarity affects other-regarding preferences in pet dogs. Scientific Reports. 5.

19. Somppi, S., Törnqvist, H., Hänninen, L., Krause, C., & Vainio, O. (2012). Dogs do look at images: eye tracking in canine cognition research. Animal cognition, 15(2), 163-174.
20. Teglas, E., Gergely, A., Kupan, K., Miklosi, A and Topal, J. (2012) Dogs' gaze following is tuned into human communicative signals. Curr biology. 7; 22 (3) 209-212.
21. Vaataja, H and Pesonen,E.K. (2013) Ethical issues and guidelines when conducting HCI studies with animals.CHI'13 Extended Abstracts on Human Factors in Computing
22. Williams, F. J., Mills, D. S., & Guo, K. (2011). Development of a head-mounted, eye-tracking system for dogs. Journal of neuroscience methods, 194(2), 259-265.
23. Weilenmann, A. and Juhlin, O. (2011) Understanding people and animals: the use of a positioning system in ordinary human-canine interaction. In Proc. CHI '11. ACM. 2631-2640
24. Zamansky, A., Baskin, S. and Anavi-Goffer,S.(2015) Digital Game Design for Canines: Getting to know your users?. BHCI'15. Animal Computer Interaction Workshop.

Towards a Wearer-Centred Framework for Animal Biotelemetry

P. Paci, C. Mancini and B.A. Price

Department of Computing and Communications, The Open University, Milton Keynes, UK.
patrizia.paci@open.ac.uk

Introduction

The emerging discipline of Animal-Computer Interaction (ACI) aims to understand the relation between animals and technology in naturalistic settings, to design technology that can support animals in different contexts and to develop user-centred research methods and frameworks that enable animals to take part in the design process as legitimate contributors [11]. Given existing interspecies differences and communication barriers, measuring the behaviour of animals involved in ACI research can be instrumental to achieving any or all of these aims, as a way of gauging the animals' patterns, needs and preferences. Indeed, measuring behaviour is a common practice among ACI researchers, who take various approaches to this task [5,15,17,24]. In this respect, the use of biotelemetry devices such as VHF tags and GPS trackers, or bio-logging and environmental sensors has a significant potential [22].

At the same time, biotelemetry has been used for many years in many areas of biological research. Biotelemetry is used to improve the quality of physiological and behavioural data collected from animals and in an attempt to reduce researchers' intrusion in the animals' habitat [2]. However, there is evidence that carrying biotelemetry tags may influence the bearer's physiology and behaviour [20]. Such impacts interfere with the validity of recorded data [14] and the welfare of individual animal wearers [1,3,13]. Neither of these effects are compatible with the animal-centred perspective advocated by ACI, on both scientific and ethical grounds. Our analysis of current body-attached device design and biotelemetry-enabled studies points to a general lack of wearer-centred perspective. To address these issues, we have developed a framework to inform the design of wearer-centred biotelemetry interventions, in order to support the implementation of animal-centred research methodologies and design solutions in ACI and other disciplines.

Background

Tracking movements and measuring vital parameters remotely in animals through biotelemetry tags has allowed the acquisition of ecological, physiological and behavioural information usually inaccessible with other observational methodologies [2]. Since the 1960s, body-attached devices have been used for the study of targeted species in their natural environment, the refinement of experimental procedures, and the implementation of biodiversity conservation strategies [2,18]. Although the use of biotelemetry has revolutionized data-gathering practices in animal biology and ecology, side effects caused to the wearer by the device itself can interfere with the phenomena that are being monitored, thus producing unreliable and biased data. For one example, when studying the foraging behaviour of penguins using transmitters, the tag attached on the body can increase drag, thus reducing the swimming speed and altering the very hunting patterns being investigated [23]. Impacts on individuals can be physically manifested, such as those occurring on the site of the attachment (e.g. fur abrasion, limb swelling, or wounds), or less obviously perceivable to researchers, such as alterations of physiological parameters (e.g. variations in the metabolic activity and body temperature). Changes in the normal behaviour can be apparently unrelated with the presence of the device (e.g. decrease in foraging behaviour) or clearly derived from it (e.g. abnormal grooming in the attempt of removing the foreign body) [9].

Therefore, although measuring behaviour can play a key role in ACI and other research as a way of understanding otherwise non-accessible aspects of animals' habits, needs and preferences, the use of biotelemetry can have implications for the scientific validity of recorded data as well as serious consequences for the welfare of the animals involved in research procedures. To reduce such device-induced impacts, animal

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

welfare scientists have proposed guidelines and recommendations for improving the experimental designs of studies that employ biotelemetry. Namely, they have pointed out the need to re-design both body-attachment methods and tag features, such as weight, shape and colour in a way that better conforms to the wearer [1,3,13,23]. Arguably, whereby negative effects induced by a device are minimized or removed, the quality of collected data can be improved as well as the welfare of the animal being monitored.

However, although in principle such guidelines aim to bring the perspective of the animal to the attention of researchers and designers, when it comes to application details they often lack the very perspective they advocate. For example, in one of their recommendations, researchers discourage the use of the red hue in device components, suggesting that this particular colour can be interpreted as blood by predators or conspecifics [3]. Indeed, this may be the case if said predators or conspecifics are able to see colours in the same way that humans do, and more importantly, if they use sight as the guiding sense towards blood, and colour as blood's characterising feature. However, many mammal species have di-chromatic vision [4] and many such predators are driven to prey by scent rather than sight. For example, wolves have a highly sophisticated olfactory system they use to track prey, but a scarce ability to detect red objects, perceiving them in shades of grey instead [12]. We agree that a red harness or tag encase could generate an impact (for example, by disrupting the camouflage of an animal). However, we argue that design recommendations should be informed by criteria that systematically extend beyond the human perspective (e.g. associating the colour red with blood and colour as a salient marker of blood). However, to date there is no design framework that can help researchers and designers to systematically account for and reconcile the often diverging perspectives of the animal wearer and of the human user, and the design requirements that derive from both. For example, ecologists often use coloured tags for marking the animals they are studying because they need to easily identify individuals during field observations; but this can be detrimental for the animals if they become easily detectable to ill-intentioned humans, potential predators or prey. To address this gap, we have developed a *wearer-centred* framework with the intent of bringing the wearers' perspective into the process of designing biotelemetry devices.

Designing for wearability

In Interaction Design [16] it has long been established that good interactions are designed 'around' users - their characteristics, those of their activities and those of their environment – systematically informed by established design principle (e.g. perceivability, affordance). The same user-centred perspective characterises designers' approach to novel forms of interaction enabled by ubiquitous computing technology, whether these are intentional or unintentional, explicit or implicit [7]. Although these include physical interactions resulting from the use of, or contact with, wearable technology, a design framework to support the design of wearer centred devices is still lacking. Developing such a framework is our aim, particularly with reference to the design of biotelemetry for animals, in order to improve their *wearer experience* and therefore reduce the negative aspects related to the device presence.

Providing good user experience is a main goal of user-centred technology, with the fundamental assumption that the technology the user interacts with is directly relevant to them and their activities. But what is the equivalent of a 'good experience' when the technology one interacts with is not directly relevant to one's intentions and activities? This is arguably the position in which animal wearers of biotelemetry find themselves in when they physically interact with technology that does not serve their own purposes but those of someone else. In this case, we argue that paradoxically a good wearer experience amounts to having 'no experience' of the device: in other words, good wearable biotelemetry (for animal wearers who are not users) is one that does not get in the way of the animal's daily experiences, activities or social interactions, one that is not experienced at all. To this end, we propose that the design of wearable devices should be informed by *design principles for wearability* pertaining to the animal sensory, physical and cognitive experience, namely: *sensory imperceptibility*, *physical unobtrusiveness*, and *cognitive acceptability.* Sensory perceptibility refers to a wider range of senses in comparison with those of humans (e.g. electro-receptive animals can sense the electric fields emitted by the tag [8]) as well as to a wider spectrum of sensitivity (e.g. birds such as raptors may perceive coloured devices at a much greater distance than people do, thanks to their very acute and pigmented vision [6]). Physical obtrusiveness is linked with locomotive abilities (e.g. swimming or flying movements can be limited by a tag attached in an improper location) and environmental features (e.g. dense vegetation can impede smooth movements of instrumented animals) [9]. Cognitive unacceptability refers to the psychological condition of those animals that, being aware of the device, do not accept its presence, which can lead to the development of atypical behaviour such as stereotypes (detrimental compulsions that may arise when a wearer cannot express its natural behaviour because of the tag). We propose that, when designing tags, considering the abovementioned principles in relation to the animal's biology, and consistent with the way in which the animal may experience the device, can help ensure that the devices' features do not generate an impact on the wearer.

Furthermore, because individual animals are part of wider social ecologies, these principles do not just relate to the wearers themselves, they also apply to other individuals significantly interacting with them [20]. These *significant others* include potential prey and predators of the wearer, or conspecifics such as sexual mates or members of the same social group, whose interaction with the wearer can be significantly altered due to the device. For example, a potential mate might perceive the tag of an individual, experience it as physically obtrusive, and find it cognitively unacceptable, thus preferring non-instrumented partners instead of the tagged individual. Therefore, in order to design devices that are imperceptible, unobtrusive and acceptable, researchers need to carefully consider the sensory, physical and cognitive characteristics of the wearer, and those of their significant others, and how their environment and activities influence those characteristics.

Below we discuss an example to illustrate how our framework can be operationalized. We show how our proposed wearability principles, considered from the perspective of all stakeholders (wearers, significant others and human users) in relation to their sensory, physical, cognitive, behavioural and environmental characteristics, can inform a systematic requirements analysis. We propose that this can in turn lead to the identification and development of wearer-centred design solutions. For our illustrative example, we consider a North-America population of *red foxes* (*Vulpes vulpes*) as species of interest and *hearing* as design-informative variable for the principle of *sensory imperceptibility*. In order to be *aurally imperceptible* as not to influence critical and delicate activities (e.g. by interfering with mating calls, alerting and dispersing prey, or disrupting ambushes), a device should not produce any frequency audible by the animals instrumented [1], or by their *significant others* (i.e. their conspecifics, their prey, and their predators). This particular requirement is not incompatible with those of human users (e.g. researchers), since in this case there is no interest for ecologists in detecting instrumented animals through acoustic signals. The possibility for a biotelemetry tag of emitting ultrasounds hearable by animals have been demonstrated by studies on bat dataloggers which revealed the emission of ultrasonic bands in the measure of circa 33,000 Hz [21]. In this respect, the framework asks designers to assess the presence of detectable radiation from the device in relation with the aural capabilities of all the animal species that are likely to be involved or affected (i.e. instrumented animals plus their significant others within the geographical context and distribution area in question). In this simplification, the wearer species, foxes, have an audiogram within the approximate range of 51-48,000 Hz [10] (for comparison: humans' audiogram is commonly 20-20,000 Hz). The audiogram of foxes' significant others varies: their typical prey, mice, have an aural sensitivity of circa 1,000-91,000 Hz [19]; their potential (but not regular) predators living in the studied area, coyotes (*Canis latrans*), are likely to have an aural sensitivity of circa 67-45,000 Hz (although their exact hearing range is not known, it is likely similar to that of canids such as dogs [19]). This means that in order to meet the hearing requirement, a device used on foxes should not emit auditory signals within the frequencies of 51-91,000 Hz, which is the combined minimum and maximum frequency hearable by at least one of the three species. If the envisaged tag does not produce any noise within this range, then the audible aspects of the *imperceptibility* principle are fulfilled. On the contrary, if the technology used generates a resonance, the device design should be revised and technologies that do not produce ground noise in the 51-91,000 Hz span should be used (or designed) instead. Should this not be possible due to current technological limitations, trade-offs should be considered following a scale of importance, where importance is determined by the severity of the expected impact produced by the electronic tag on the wearer. To continue with our example, as we mentioned, coyotes are accidental fox predators while mice are regular fox quarry. Therefore the predatory impact of coyotes on vulpines is less significant than fox hunting failure on mice, which can lead to starvation, especially because mice rely on their very high sensitive hearing system to escape from predators, whereas coyotes' hunting behaviour on foxes is principally driven by smell and sight.

The same assessment process should be carried out for each known significant other (e.g. mice are not the only prey foxes rely upon) and for all the relevant variables associated with the biological and environmental characteristics of the animals in question. More specifically, these are related with the sensory, physiological, morphological and psychological characteristics of animals, their physical and social environment, their living conditions, daily activities, behaviours, and movements.

To validate our framework, we are currently designing evaluation studies comparing a range of off-the-shelf biotelemetry wearables designed for felines with the purpose of observing which tag features, and to what extent, they are liable to produce side effects on the wearer (for example, fur abrasion on the site of attachment, increment in grooming, or noise intrusiveness). The findings will serve as an initial validation of our framework. As a complementary form of validation, we will then apply the framework to inform the design of prototype tags, which we will evaluate against off-the shelf devices. The comparison between the impacts produced by the commercial products and our prototype will provide insights as to the validity of the framework. To begin with, we will work with cats as the model species of choice, with a view to extend the work to other species. Our experimental designs will be assessed and approved by the Animal Welfare and Ethical Review Body of the Open University and will also conform to the University's ACI Research Ethics Protocol.

Conclusions

Biotelemetry has played an important role in the development of behavioural science, and biological sciences more generally. Similarly, it could play a key role in ACI. However, we argue that, if biotelemetry is to truly contribute to the kind of animal-centred research that ACI aims for, the perspective from which this technology is designed needs an essential and systematic shift. We propose that our wearer-centred framework, and the solutions that can be derived from its application, can significantly increase the potential of biotelemetry in ACI, both as a part of animal-centred applications and as a methodological instrument to conduct animal-centred research.

References

1. Casper, R.M. (2009). Guidelines for the instrumentation of wild birds and mammals. Animal Behaviour 78, 1477-1483.
2. Cooke, S.J. (2008). Biotelemetry and biologging in endangered species research and animal conservation: relevance to regional, national, and IUCN red list threat assessment. Endangered Species Research 4, 165-185.
3. Hawkins, P. (2004). Bio-logging and animal welfare: practical refinements. Memoirs of National Institute of Polar Research 58, 58-68.
4. Jacobs, G.H. (2009). Evolution of colour vision in mammals. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 2957-2967.
5. Johnston-Wilder, O., Mancini, C., Aengenheister, B., Mills, J., Harris, R., Guest, C. (2015). Sensing the shape of canine responses to cancer. Proceedings of ACM Advances in Computing Entertainment 2015 (Iskandar, 16-19 November 2015).
6. Jones, M.P., Pierce Jr K.E., Ward, D. (2007). Avian vision: a review of form and function with special consideration to birds of prey. Journal of Exotic Pet Medicine 16, 69-87.
7. Ju, W., Leifer, L. (2008). The design of implicit interactions: making interactive systems less obnoxious. Design Issues 24, 72–84.
8. Keinath, J.A., Musick, J.A. (1993). Movements and diving behavior of a leatherback turtle, Dermochelys coriacea. Copeia 4, 1010-1017.
9. Kenward, R.E. (2001). A manual for wildlife radio tagging. London Academic Press, London.
10. Malkemper, E.P., Topinka, V., Burda, H. (2015). A behavioral audiogram of the red fox (Vulpes vulpes). Hearing Research 320, 30-37.
11. Mancini, C. (2011). Animal-computer interaction: a manifesto. Interactions 18, 69-73.
12. Mazis, G. (2008). The world of wolves. Environmental Philosophy 5, 69-91.
13. Morton, D.B., Hawkins, P., Bevan, R., Heath, K., Kirkwood, J., Pearce, P., Scott, L., Whelan, G., Webb, A.

(2003). Refinements in telemetry procedures: seventh report of the BVAAWF/FRAME/RSPCA/UFAW joint working group on refinement, part A. Laboratory Animals 37, 261-299.
14. Murray, D.L., Fuller, M.R. (2000). A critical review of the effects of marking on the biology of vertebrates. In: Boitani L, Fuller TK (eds) Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York pp 15-16.
15. Pons, P., Jaen, J., Catala, A. (2015). Developing a depth-based tracking system for interactive playful environments with animals. Proceedings of ACM Advances in Computing Entertainment 2015 (Iskandar, 16-19 November 2015).
16. Preece, J., Sharp, H., Rogers, Y. (2015). Interaction design: beyond Human-Computer Interaction (fourth edition). John Wiley & Sons.
17. Robinson, C., Mancini, C., van der Linden, J., Guest, C., Harris, R. (2014). Canine-centred interface design: supporting the work of diabetes alert dogs. Proceedings of ACM CHI 2014 on Human Factors in Computing Systems (Toronto, 26 April-01 May 2014), 3757-3766.
18. Rutz, C., Hays, G.C. (2009). New frontiers in biologging science. Biology Letters 5, 289-292.
19. Strain, G.M. (2003). Hearing frequency ranges for dogs and other species. <http://www.lsu.edu/deafness/HearingRange.html>. Accessed 15 March 2016.
20. Walker, K.A., Trites, A.W., Haulena, M., Weary, D.M. (2012). A review of the effects of different marking and tagging techniques on marine mammals. Wildlife Research 39, 15-30.
21. Willis, C.K.R., Jameson, J.W., Faure, P.A., Boyles, J.G., Brack, V., Cervone, T.H. (2009). Thermocron iButton and iBBat temperature dataloggers emit ultrasound. Journal of Comparative Physiology B 179, 867-874.
22. Wilmers, C.C., Nickel, B., Bryce, C.M., Smith, J.A., Wheat, R.E., Yovovich, V. (2015). The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. Ecology 96, 1741-1753.
23. Wilson, R.P., McMahon, C.R. (2006). Measuring devices on wild animals: what constitutes acceptable practice? Frontiers in Ecology and the Environment 4, 147-154.
24. Zeagler, C., Gilliland, S., Freil, L., Starner, T., Jackson M. (2014). Going to the dogs: towards an interactive touchscreen interface for working dogs. Proceedings of ACM UIST 2014 on User Interface Software and Technology (Honolulu, 5-8 October 2014), 497-507.

Playful UX for Elephants

Fiona French, Clara Mancini, Helen Sharp

Faculty of Life Sciences ad Computing, London Metropolitan University, London, UK
f.french@londonmet.ac.uk
Faculty of Mathematics, Computing and Technology, The Open University, Milton Keynes, UK
c.mancini@open.ac.uk, h.sharp@open.ac.uk

Abstract

This case study describes approaches to the challenge of designing interfaces for an elephant that enable her to control playful systems in her enclosure, for the purpose of enriching her environment. Our contribution to the symposium will showcase the progress of the enrichment toys and explain in detail how we have collected feedback during participatory design sessions with our play-tester Valli, a female Asian elephant. We have attempted to gain information about her enthusiasm for interacting with different systems and also establish how effectively she can use different interfaces by measuring her responses during the sessions.

Background

There is a consensus among researchers [1] [4] that offering animals more control is a positive step, as it provides an opportunity for them to seek out and respond to a positive stimulus. Mills et al [2] explain the importance for animals of being able to anticipate and prepare for what is expected, as well as being able to respond spontaneously to changes in the environment, in order to maintain an optimal physiological and psychological condition. At the same time, the authors highlight the importance of interventions that can enable animals to enrich their cognitive experience through positive stimuli.

Our aim for this work is to develop some novel forms of cognitive and sensory enrichment for captive elephants, facilitated through the use of technology. Mancini et al [3] emphasise the opportunities that technology affords, providing the means to offer personalised experiences to individual captive animals by monitoring usage and implementing smart, adaptive systems. Conditions for captive elephants in the UK vary with regard to their environmental conditions, their social opportunities and the degree of human interaction they experience, amongst other factors. This highlights the potential utility of a system that is adaptable to suit an individual animal in a specific context.

Ultimately, our goal is to design some interactive toys that encourage playful behaviour, which is widely regarded as a hallmark of good welfare in captive animals [1] [5], indicating that they are not stressed and are therefore able to cope, not only with their environmental conditions, but also with the uncertainty inherent in a playful situation.

Smart toys can be designed to provide the animals with both control and stimulation appropriate to their needs. Specifically, on the one hand, the interaction with smart toys can enable animals to access a range of stimuli in the form of system output; on the other hand, they can afford the animals control over what stimuli to access and when to do so by expressing their choices in the form of system input. Both forms of engagement have significant potential for cognitive enrichment.

Work to date

Our current participant, Valli, was orphaned at birth and has been living with human companions at Skanda Vale Ashram [6] for over thirty years. She and her keepers are the main stake-holders, although we anticipate playful systems being tried in various zoos in UK and Ireland.

Figure 1: Elephant shower

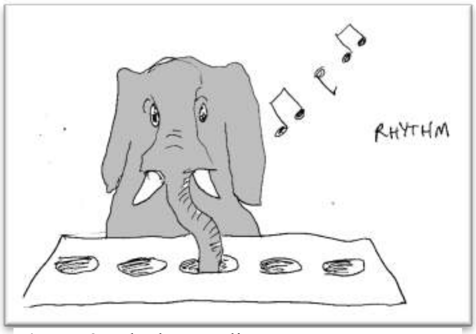


Figure 2: Elephant radio

Figure 3: Elephant theramin

Some of the concepts being explored include: (i) a set of shower controls, enabling Valli to control water supply, quality of spray and possibly temperature; (ii) an elephant "radio" system, enabling her to select between different audio outputs; (iii) an acoustic instrument that enables her to manipulate the quality of the sounds emitted.

These toys are works-in-progress, and are being developed using participatory design as much as possible. The design team includes the developer, the keepers and the target user – Valli. All of these stake-holders bring different perspectives to the design challenge, which raises some questions about how we obtain useful feedback. For example, whereas a human might be able to offer a judgment about which controls are most effective for regulating a shower, an elephant can only show us her preferences through her actions.

Our early studies have therefore focused mainly on testing different input mechanisms in order to discover which have most intrinsic appeal for Valli. We have judged a control to be successful if she investigates it without a food reward and is able to easily activate it.

Process

For every design session there are four clear goals that provide focus to the exercise for both the developer and the elephant keepers:

1. Enrichment Goal – e.g. Investigate acoustic enrichment: Does Valli show interest in low frequency audio samples and if so, in what range?
2. Playfulness Goal – e.g. Identify objects and modalities that are conducive to play: Does Valli show signs of object or locomotor play when interacting with this system? Is Valli motivated to use this control in the absence of a food reward?
3. Usability Goal – e.g. Test functionality and evaluate usability: Can Valli effectively use her trunk to activate this switch? Does Valli understand the connection between her input to the system and the resulting output?
4. Technical Goal – e.g. Construct a system that uses capacitance sensing to detect trunk proximity.

At present, during each session, the keepers provide feedback on the enrichment and playfulness potential of a device, while we monitor the extent to which the usability and technical goals are achieved.

We have tested some prototype interfaces, paying particular attention to the aesthetics of the feedback, as we are interested in the quality of the experience for our users. We are aiming to provide naturalistic controls that encourage experimentation and investigation, using natural trunk movements. This is because elephants normally use their trunks to interact with the world and explore places that might not be visible, acquiring chemical and tactile information [7].

Behavioural measurements are used in the design process in two distinct ways: (i) as a way of enabling Valli to control the system (i.e. enabling her to give the system an input) and obtain a stimulus (i.e. receiving an output in return); (ii) as a way for us to understand what kinds of systems might have elephant appeal.

Figure 4: Valli investigates pipe buttons with hidden sensors

With regards to the first kind of behavioural measurement, her interaction with the most successful controls (from a usability perspective) has been facilitated using hidden sensors that can detect movement or measure proximity. The controls have inbuilt feedback mechanisms, activated by the sensors. This means they provide a tactile indication that she has approached or activated a control, which she is able to sense using her trunk tip. For practical reasons, a control system might need to be mounted in a place where she cannot see the buttons (for example through a hole in the wall), and therefore offering feedback that reinforces her intention and behaviour can be very important.

The second type of behavioural measurement involves collecting data about how the system is used. As we are not sure what kinds of sensory output an elephant would necessarily enjoy, the more choices we can offer Valli, the better. To understand Valli's preferences, the usability and playfulness afforded by the different options are being assessed according to how she interacts with a particular system, noting how often and for how long interactions take place. This way, we hope to be able to transcend the (inevitably) human-centred aspect of the designs and allow the elephant to offer us valuable feedback through her selections and modes of interaction (e.g. when, how and how long for she engages with a prototype). This philosophy underpins our approach to participatory design – we enable Valli to make choices about the systems that then contribute to the iterative design process.

Figure 5: Button made from sewing machine pedal

As an example, we have been observing Valli's engagement with buttons made from a variety of materials, including one made using a sewing machine pedal. Watching her spend several minutes exploring a wooden button frame but continuously fail to push the embedded hard plastic pedal led to a revision of the design concept. The keepers concluded that Valli could easily learn to push a large sewing machine pedal if she had sufficient motivation. However, as well as avoiding any associations with food (used to help with training), we aim to develop a system that has intrinsic interest for an elephant. The new design uses natural materials (textiles) in a button that requires no pressure from Valli's trunk but never-the-less provides haptic feedback in the form of patterns of vibration [8].

Future plans

The next stage is to determine what kinds of outputs Valli chooses when she understands that she is in control of the selection. We are investigating systems with acoustic and haptic feedback, as well as offering control over environmental features such as water supply. We will offer Valli a series of identical buttons with similar vibrotactile interfaces. The buttons will be distinguished by the kinds of feedback they activate and we will collect data about how much she uses each button.

In the future, we hope to move from using a digital on/off system for controlling discrete outputs (for example, sound samples pre-selected by a human) to a system that can capture more detailed data about Valli's preferences.

For an acoustic toy, we plan to develop an instrument (similar to a theramin) that allows the user to have control over the quality of the output, for example by manipulating volume, pitch or timbre. We will collect data from the toy so that we can monitor Valli's reactions in the absence of keepers and visitors, hoping to discover what kinds of acoustic experiences engage her interest over a longer period of time. Any results will be fed back into the development process, to refine the design of the toy so that it meets Valli's needs. This kind of data could also be useful for studies that focus on measuring the welfare impact of a technological intervention. There are a number of possible design solutions for such a toy and these will be presented during the symposium with the results of initial trials.

Ethical statement

The Open University Animal Welfare Ethical Review Body (AWERB) have approved the research protocol for this non-licensed research involving elephants at Skanda Vale Ashram and Blair Drummond Safari Park. The work also conforms to the Open University's ACI Research Ethics Protocol.

References

1. Young, R.J., (2003). Environmental enrichment for captive animals, Oxford, UK ; Malden, MA: Blackwell Science.
2. Mills, D.S., Dube, M.B. & Zulch, H., (2012). Stress and pheromonatherapy in small animal clinical behaviour, Chichester, West Sussex ; Ames, IA: Wiley-Blackwell
3. Mancini, C., van der Linden, J., Kortuem, G., Dewsbury, G., Mills, D. and Boyden, P. (2014). UbiComp for animal welfare: envisioning smart environments for kenneled dogs. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14). ACM, New York, NY, USA, 117-128. DOI=http://dx.doi.org/10.1145/2632048.2632073
4. Buchanan-Smith, H.M. & Badihi, I., (2012). The psychology of control: Effects of control over supplementary light on welfare of marmosets. Applied Animal Behaviour Science, 137(3-4), pp.166–174.
5. Oliveira, A.F.S. et al., (2010). Play behaviour in nonhuman animals and the animal welfare issue. Journal of Ethology, 28(1), pp.1–5.
6. Skanda Vale Ashram: http://www.skandavale.org/
7. Plotnik, J.M & de Waal (2014). Thinking with their trunks: elephants use smell but not sound to locate food and exclude nonrewarding alternatives. Animal Behaviour, 88, pp.91–98.
8. French, F., Mancini, C. & Sharp, H. (2016) Trunk-enabled toys. (future publication) Part of HCI goes to the Zoo Symposium at CHI 2016

.

Animal-Computer Interaction (ACI): An Analysis, a Perspective, and Guidelines

Egon L. van den Broek

Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
vandenbroek@acm.org

Abstract

Animal-Computer Interaction (ACI)'s founding elements are discussed in relation to its overarching discipline Human-Computer Interaction (HCI). Its basic dimensions are identified: agent, computing machinery, and interaction, and their levels of processing: perceptual, cognitive, and affective. Subsequently, three seminal studies are discussed, the ACI community should be become acquainted with. Next, three guidelines are defined that could help ACI to gain further maturity. We close with a brief conclusion.

Introduction

The previous two years, the first two International Congresses on Animal-(Human-)Computer Interaction (ACI) have been organized[2]. As such, a new vibrant subfield of HCI has emerged [10], which is sustained with various follow-up initiatives, including the current symposium. However, since its origin HCI itself has always been a research field on the move, lacking a single, stable, generally accepted definition. As such, HCI provides a fragile foundation for ACI. Recently, Carroll [4] gave the following definition: "*Today, HCI is a vast and multifaceted community, bound by the evolving concept of usability, and the integrating commitment to value human activity and experience as the primary driver in technology.*"

Directly derived from HCI and ACI's names, three basic dimensions can be identified:

- agent
- computing machinery
- interaction

and, subsequently, the following levels of processing can be distinguished:

i) intra (e.g., attention / uptime):
   a. low level perceptual (e.g., pattern recognition [2,18]);
   b. high level cognitive (e.g., models [19,20]); and
   c. affective (e.g., emotions), which interacts with all other levels [6,7,15,16];
ii) inter (e.g., as in social media [9] and cooperative annotation [20]); and
iii) role-based (e.g., functioning within an organization [17]),

where with the latter two a network perspective should be taken, the first process concerns one agent. Each of these three levels hold for both agents and machinery, except for affective processing. As such, HCI and ACI can increase our understanding of both its agents (e.g., via computational models) and computing machinery as well as their interaction, which should be considered as ACI's primary focus [10]. Note that, although generally assumed, the agent does not have to be human or animal, it can be any agent, including even computing machinery (e.g., a robot [14]).

Although the decomposition in three levels of processing is valid for any agent, in the vast majority of cases the agents have been humans. Moreover, animals differ from human users on each of these levels and animal species have far from homogeneous characteristics [10]. So, knowledge transfer from HCI to ACI should be considered

[2] The ACI URL: http://animalcomputerinteraction.org [Accessed 25 March 2016].

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

with the utmost care! Starting directly from this concern, this article will describe initial ideas to extent Mancini's research agenda [10], consisting of two components: i) some studies that deserve attention from the ACI community (and seem to be overlooked, so far) and ii) some concrete initial suggestions for a methodological foundation for ACI research. Last, a brief conclusion is provided.

Some … studies that deserve some attention

This section presents by no means an exhaustive survey. It should be considered as a sample from scientific literature highly relevant for ACI; but, perhaps, overlooked if not discussed. Three scientific articles from the last 25 years to 10 years. For reasons of brevity, only a triplet of studies is discussed, many, many other seminal studies can be considered as at least equally important for ACI (e.g., Harlow's work on "the nature of love" [14]) and, more recently, the work of Van Eck and Lamers, who studied whether or not it is possible to play computer games against animals [22].

ACI is a very interesting idea, not only from a HCI perspective; but, also, from an Artificial Intelligence (AI) perspective. In 2011, Mancini [10] stated *"... quite literally, the elephant in the room of user-computer interaction research. The time has come to acknowledge the elephant, to start talking about ACI as a discipline in its own right, and to start working toward its systematic development."* 21 years before, Rodney A. Brooks already noted that "Elephants don't play chess" [3]: "*There is an alternative route to Artificial Intelligence that diverges from the directions pursued under that banner for the last thirty some years. The traditional approach has emphasized the abstract manipulation of symbols, whose grounding in physical reality has rarely been achieved. We explore a research methodology which emphasizes ongoing physical interaction with the environment as the primary source of constraint on the design of intelligent systems.*" His claim helps us to identify three important notions for ACI, namely:

1. Brooks [3] shows to be an advocate for empirical computer science, more specifically: empirical AI. In particular, he stresses the importance of the interaction of agents (e.g., a human, animal, and robot) with each other and their environment.
2. He also explicitly distinguishes two branches of AI, that process information on entirely other levels: symbols and signals. This is very interesting as both levels are of importance for ACI, although the latter seems to outrun the former. The symbolic level is considered as typically and even exclusively linked to human intelligence.
3. Throughout its complete existence, AI has struggled in its attempts to fulfill its promises. Implicitly, Brooks seems to suggest that AI has taken the wrong turn or, at least, should divide its attention. Up to this day, the former is a topic of debate. However, there is a general agreement on that at least the latter is needed. From that perspective, ACI should perhaps even be preferred over HCI. Humans claim to be much more advanced than animals should make it simpler to understand the latter than the former and, hence, should make it simpler to understand ACI than HCI. As such, ACI could be a model for HCI, as is the case with many other sciences, where specific animals (e.g., rats and monkeys) serve as a model for humans or robots (cf. [1]).

In the 80s, an interesting HCI initiative was implemented at the MIT Media Lab [5]. It concerned an artificial animal they named Noobie (short for "New Beast") that interacted with children. Noobie looked like furry Sesame Street's Muppet (cf. Sesame Street's Pino). So, it is HCI or human-animal interaction and no ACI. Druin [5] concisely describes Alan Kay perspective on animal models: "*the use of animal agents is a wonderful way of stretching what we already know about representing animal behaviors, as well as what we know about rendering and animating figures. Therefore, if we are able to give children the tools to explore animal behavior, then we as tool-makers may learn just as much about our tools as about animal behavior.*" This is on par with what ACI aims: continuously learn about animals, adapt the animal models embedded in the computer, and, consequently, improve the actual ACI [10]. In other words, there are many possible paths that can be followed to realize and improve on ACI. Noobie's team took five evaluation criteria, which can be adopted by ACI as well [5]:

1. Comprehension: How much time did it take the animal to figure out what to do with the computer?
2. Ease of use: How easy was it to use the computer?
3. Interaction styles: How does the animal interact with the computer?
4. Attention span: How much time did it take before the animal lost interest?
5. Expectations: What did the animal want or expect to do with the computer (if anything at all)?

In 2006, Kerepesi et al. [8] reported on a study that is highly relevant for ACI. They compared human-animal (dog) and human-robot (AIBO) interactions. So, as with [5], no ACI; but, nevertheless, work that is highly relevant for ACI. Let us mention three concerns addressed in this article that are of interest to the ACI community:

1. The importance of interaction. Kerepesi et al. [6] conclude "*... that more attention should be paid in the future to the robots' ability to engage in cooperative interaction with humans.*" So, again, as with the other two articles, interaction is posed central in understanding the agents involved and their relations.
2. Levels of complexity of the agent and the computing machinery. In addition to Kerepesi et al.'s work [8], a few years ago a vivid discussion on robot nannies emerged [14]. Although it concerned human children and, hence, involved HCI or, more specifically, human-robot interaction, it can be conceived as an interesting intermediate step between HCI and ACI. In some aspects, children are closer to some animals (e.g., apes and, even, monkeys) than they are to adult humans.
3. Study temporal structures. Kerepesi et al. [8] stress the importance of detecting temporal patterns (or T-patterns). They conclude that *"whether humans were playing with dog or AIBO had a significant effect on the structure of the patterns. Both children and adults terminated T-patterns more frequently when playing with AIBO than when playing with the dog puppy, which suggest that the robot has a limited ability to engage in temporally structured behavioural interactions with humans. As other human studies suggest that the temporal complexity of the interaction is good measure of the partner's attitude ..."* (e.g., cf. [15]).

Taken together, ACI would benefit from historical reflection (cf. [16]). Next to HCI literature, knowledge from biology, psychology, communication science, AI, and robotics should be reviewed. Moreover, ACI articles should take a clear position in how they want to contribute: i) The animal can be a model of humans and as such help us to understand humans, ii) ACI can learn us about the animal in a semi-controlled manner, iii) ACI can also learn us about the computer's pros and cons, and, last, iv) it can learn about the interaction between animals and computers.

Some … suggestions

To conduct valid, replicable studies, ACI needs to meet some guidelines or as Mancini baptized it: a research agenda [10]. This section provides three of such guidelines, which is by no means an exhaustive list; but, perhaps, it could serve as starting point. These guidelines link directly to Mancini research agenda [10], extending this agenda. These guidelines are derived from a set of guidelines introduced in the context of affective computing [15].

Context

When animals and humans interact with each other, they are able to use implicit situational information, or *context*, to increase the conversational bandwidth. Unfortunately, this ability to convey ideas does not transfer well to animals interacting with computers. In traditional interactive computing, human users have an impoverished mechanism for providing input to computers. Consequently, computers are currently not enabled to take full advantage of the context of the animal-computer dialogue. By improving the computer's access to context, we increase the richness of communication in ACI and make it possible to produce more useful computational services (cf. [7]).

As is stated above, capturing context is easier said than done. Handling context is even considered as one of AI's traditional struggles [15]. Perhaps this can be attributed partly to the fact that in the vast majority of cases, research on context aware computing has taken a technology-centered perspective as opposed to an animal or human-centered perspective. This technology push has been fruitful though, among many other techniques, sensor networks, body area networks, GPS, and RFID have been developed [2,4,6,7,14,15]. Their usage can be considered as a first step towards context aware computing. However, not only the gathering is challenging but also processing (e.g., feature extraction) and interpretation are hard [15,16].

Potentially, context aware computing can aid ACI significantly. Biosensors can be embedded in jewelry (e.g., a belt or necklace), in consumer electronics (e.g., a cell phone or music player), or otherwise as wearables (e.g., embedded in cloths or as part of a body area network). Connected to (more powerful) processing units they can record, tag, and interpret events and, in parallel, tap into animals' reactions through our physiological responses.

Validation

In the pursuit to study animal behavior in a more or less controlled manner, a range of methods have been applied: multimedia (e.g., images, audio, and multimedia), games, agents, and real world experiences [15]. However, how to know what these methods actually triggered with the animals studied? This is a typical concern of validity, which is a crucial issue for ACI. Validity can be obtained through various approaches. Here, we will discuss four of them [15].

Content validity refers to the degree …

- of expert's agreement on the domain of interest (e.g., limited to a specific application or group of animals);
- to which a feature (or its parameters) of a given signal represents a construct; and
- to which a set of features (or their parameters) of a given set of signals represents all facets of the domain.

Criteria-related validity handles the quality of the translation from the preferred measurement to an alternative, rather than to what extend the measurement represents a construct. Interactions are preferably measured at the moment they occur; however, measurements before (predictive) or after (postdictive) the particular event are sometimes more feasible. The quality of these translations is referred to as predictive or postdictive validity. A third form of criteria-related validity is concurrent validity: a metric for the reliability of measurements applied in relation to the preferred standard. For instance, the more types of interaction are discriminated the higher the concurrent validity.

A *construct validation* process aims to develop a nomological network (i.e., a ground truth) or an ontology or semantic network, build around the construct of interest. Such a network requires theoretically grounded, observable, operational definitions of all constructs and the relations between them. Such a network aims to provide a verifiable theoretical framework. The lack of such a network is one of the most pregnant problems ACI is coping with. Often a statement such as "*The term animal(s) is loosely used throughout to refer to nonhuman animals.*" [10] is made to cover this[1]. However, better would be: "*any of a kingdom (Animalia) of living things including many-celled organisms and often many of the single-celled ones (as protozoans) that typically differ from plants in having cells without cellulose walls, in lacking chlorophyll and the capacity for photosynthesis, in requiring more complex food materials (as proteins), in being organized to a greater degree of complexity, and in having the capacity for spontaneous movement and rapid motor responses to stimulation.*" [11].

Ecological validity refers to the influence of the context on measurements. We identify two issues:

- Natural ACI events are sparse, which makes it hard to let animals cycle through a range of interactions in a limited time frame; and
- The interactions that occur are easily contaminated by contextual factors; so, using a similar context as the intended ACI application for initial learning is of vital importance.

From a measurement-feasibility perspective, an easy way out would be to conduct interaction measurements in controlled laboratory settings. However, even this by itself could be a stressor for animals. Moreover, it makes results poorly generalizable to real-world applications.

Triangulation

Triangulation is the strategy of combining multiple data sources, investigators, methodological approaches, theoretical perspectives, or analytical methods within the same study [9]. In the operationalization, this provides methodological instruments to separate the construct under consideration from irrelevancies. We propose to adopt this principle of triangulation, as applied in social sciences for ACI.

Five types of triangulation can be distinguished [12], namely:

1. *Data*: Three dimensions in data sources can be distinguished: time, space (or setting), and the recorder. Time triangulation can be applied when data is collected at different times. In general, variance in events, situations, times, places, and persons are considered as sources of noise. Extrapolations on multiple data sets can provide more certainty in such cases and anomalies can be detected and corrected.
2. *Investigator*: Multiple observers, interviewers, coders, or data analysts can participate in the study. Agreement among these researchers, without prior discussion or collaboration with one another, increases the credibility to the observations. Par excellence, this type of triangulation can be employed on including context and unveiling events.
3. *Methodological*: It can refer to either data collection methods or research designs. Its major advantage is that deficiencies and biases that stem from a single method can be countered.
   Multiple data sets (e.g., both qualitative and quantitative) and signal processing techniques (e.g., in the time and spectral domain) can be employed. Moreover, multiple features extraction paradigms, feature reduction algorithms, and classification schemes can be employed. Further, note that methodological triangulation is also called multi-method, mixed-method, and methods triangulation.
4. *Theoretical*: Employing multiple theoretical frameworks when examining a phenomenon.
5. *Analytical*: The combination of multiple methods or classification methods to analyze data. This facilitates (cross) validation of data sources.

In general, we advise to record at least 3 signals for each construct under investigation. In ambulatory, real-world ACI research much more noise will be recorded. To ensure this noise can be canceled out, we advise to record even more signals. As a rule of thumb for ambulatory research we advise to record as many signals possible, without that they interfere with the animal's natural behavior. However, a disadvantage accompanies this advice, as "*a `more is better' mentality may result in diluting the possible effectiveness of triangulation*" (p. 256) [12]. Moreover, qualitative and subjective measures should always accompany the signals (e.g., video-based ethnography). On the one hand, with animals, the use of many traditional HCI measures (e.g., questionnaires and cooperative annotations [22]) is challenging, if possible at all. On the other hand, Rhesus monkeys, pigs, and rats have already been successfully trained to use a joystick and perform computerized game-like tasks [22]. This calls for a base of biology's empirical methods for ACI, in addition to psychology's methods, as are often used in HCI.

Conclusion

ACI is interesting and a research field full of promises. However, it also requires a true interdisciplinary approach, which makes it vulnerable for criticism from all of its founding disciplines, as was discussed in the introduction (cf. [10]). This requires an interdisciplinary historical reflection and, consequently, identification of crucial scientific notions, which we perhaps all know; but, sometimes forget to apply [14-16]. Moreover, guidelines on ethical issues should be developed [14]. Väätäjä and Pesonen [13] provided an excellent starting point for this. By no means this article can be considered as an exhaustive treatment of the subject; but, perhaps, it can serve as that one more building block for a solid foundation for ACI. And this is very much needed, as undoubtedly ACI's progress is accelerating. ACI is here to stay and is heading towards a bright future, quickly emerging throughout a plethora of societal contexts.

Acknowledgments

The author wishes to express his gratitude to the three reviewers for their detailed and constructive comments. Their comments helped to improve this article significantly.

References

1. Anderson, T.L. and Donath, M. (1990). Animal behavior as a paradigm for developing robot autonomy. *Robotics and Autonomous Systems* 6(1-2), 145-168.
2. Bozkurt, A., Roberts, D.L., Sherman, B.L., Brugarolas, R., Mealin, S., Majikes, J., Yang, P., Loftin, R. (2014). Toward cyber-enhanced working dogs for search and rescue. *IEEE Intelligent Systems* 29(6), 32-39.
3. Brooks, R.A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems* 6(1-2), 3-15.
4. Carroll, J.M. (2016). Human-Computer Interaction: Brief intro. In M. Soegaard and R.F. Dam (Eds.), *The Encyclopedia of Human-Computer Interaction (2nd ed.), Chapter 2*. Aarhus C., Denmark: The Interaction-Design.org Foundation. Accessed 25 March 2016.
5. Druin, A. (1988). Noobie: The animal design playstation. *SIGCHI Bulletin* 20(1), 45-53.
6. Janssen, J.H., Tacken, P., De Vries, G.-J., Van den Broek, E.L., Westerink, J.H.D.M., Haselager, P., IJsselsteijn, W.A. (2013). Machines outperform lay persons in recognizing emotions elicited by autobiographical recollection. *Human-Computer Interaction* 28(6), 479-517.
7. Janssen, J.H. Van den Broek, E.L., Westerink, J.H.D.M. (2009). Personalized affective music player. In J.F. Cohn, A. Nijholt, M. Pantic (Eds.), *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction (ACII2009)*, 472-477. Los Alamitos, CA, USA: IEEE Computer Society.
8. Kerepesi, A., Kubinyi, E., Jonsson, G.K., Magnusson, M.S., and Miklósi, Á. (2006). Behavioural comparison of human-animal (dog) and human-robot (AIBO) interactions. *Behavioural Processes* 73(1), 92-99.
9. Koot, G., Huis in 't Veld, M.A.A., Hendricksen, J., Kaptein, R., De Vries, A., Van den Broek, E.L. (2014). Foraging Online Social Networks. In M. den Hengst, M. Israël, D. Zeng, C. Veenman, and A. Wang (Eds.), *Proceedings of the 2014 IEEE Joint Intelligence and Security Informatics Conference (JISIC2014)*, p. 312-315. Los Alamitos, CA, USA: IEEE Computer Society.
10. Mancini, C. (2011). Animal-Computer Interaction: A manifesto. *Interactions* 18(3), 69-73.
11. Merriam Webster, Incorporated (2016). Merriam-Webster Online. URL: http://www2.merriam-webster.com/cgi-bin/mwdictyns?book=Dictionary&va=animals . Accessed 25 March 2016.
12. Thurmond, V.A. (2001). The point of triangulation. *Journal of Nursing Scholarship 33(3)*, 253-258.
13. Väätäjä, H., Pesonen, E. (2013). Ethical issues and guidelines when conducting HCI studies with animals. In W.E. Mackay, S. Brewster, and S. Bødker (Eds.), *Proceedings of CHI'13 Extended Abstracts on Human Factors in Computing Systems*, p. 2159-2168. New York, NY, USA: ACM.
14. Van den Broek, E.L. (2010). Robot nannies: Future or fiction? *Interaction Studies* 11(2), 274-282.
15. Van den Broek, E.L. (2011). *Affective Signal Processing (ASP): Unravelling the mystery of emotions.* CTIT Ph.D.-thesis series No. 11-204, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.
16. Van den Broek, E.L. (2012). Affective computing: A reverence for a century of research. In A. Esposito, A.M. Esposito, A. Vinciarelli, R. Hoffmann, and V.C. Müller (Eds.), *Cognitive Behavioural Systems*, p. 434-448. Lecture Notes in Computer Science, volume 7403. Berlin/Heidelberg, Germany: Springer-Verlag.
17. Van den Broek, E.L., Jonker, C.M. Sharpanskykh, A., Treur, J., Yolum, P. (2006). Formal modeling and analysis of organizations. In O. Boissier, J. Padget, V. Dignum, G. Lindemann, E. Matson, S. Ossowski, J.S. Sichman, and J. Vázquez-Salceda (Eds.), *Coordination, Organizations, Institutions, and Norms in Multi-*

Agent Systems, p. 18-34. Lecture Notes in Computer Science, volume 3913. Berlin/Heidelberg, Germany: Springer Verlag.

18. Van den Broek, E.L., Schouten, Th.E., Kisters, P.M.F. (2008). Modeling human color categorization. *Pattern Recognition Letters* 29(8), 1136-1144.
19. Van den Broek, E.L. Kisters, P.M.F., Vuurpijl, L.G. (2005). Content-Based Image Retrieval benchmarking: Utilizing color categories and color distributions. *Journal of Imaging Science and Technology* 49(3), 293-301.
20. Van der Sluis, F., van den Broek, E.L., Glassey, R.J., van Dijk, E.M.A.G., de Jong, F.M.G. (2014). When complexity becomes interesting. *Journal of the American Society for Information Science and Technology* 65(7), 1478–1500.
21. Van Eck, W., Lamers, M.H. (2006). Animal controlled computer games: Playing Pac-Man against real crickets. In R. Harper, M. Rauterberg, and M. Combetto (Eds.), *Entertainment Computing - ICEC 2006*, p. 31-36. Lecture Notes in Computer Science, volume 4161. Berlin/Heidelberg, Germany: Springer Verlag.
22. Vuurpijl, L., Schomaker, L., Van den Broek, E.L. (2002). Vind(x): Using the user through cooperative annotation. In A.D. Williams (Ed.), *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02)*, 221-226. Los Alamitos, CA, USA: IEEE Computer Society.

Animal-Computer Interaction: Animal-Centred, Participatory, and Playful Design

Patricia Pons[1], Ilyena Hirskyj-Douglas[2], Anton Nijholt[3] and Adrian D. Cheok[4]

[1]ISSI Research Group, Department of Information Systems and Computation, Polytechnic University of Valencia, Valencia, Spain, ppons@dsic.upv.es
[2]Department of Computing, University of Central Lancashire, Lancashire, England, ihirskyj-douglas@uclan.ac.uk
[3]Human Media Interaction, University of Twente, Enschede, The Netherlands and Imagineering Institute, Iskandar, Malaysia, a.nijholt@utwente.nl
[4]Imagineering Institute, Iskandar, Malaysia and City University, London, UK, adrian@imagineeringinstitute.org

Introduction

This introductory paper presents the field of Animal Computer Interaction (ACI) and the work that is being presented at the ACI Symposium at Measuring Behavior. In recent years there has been growing interest in developing systems to improve animal's wellbeing and to support the interaction of animals within the digital domain. The field of Animal-Computer Interaction considers animals as the end-users of the technology being developed for them, orienting this process by the needs and characteristics of its non-human stakeholders. Although animals have interacted with technology long before the emergence of the ACI field [12, 18], usually their behaviors were typically anthropomorphized with their needs not being fully understood. Instead, ACI research adopts an animal-centred approach and aims to understand technologically mediated interaction of animals from their perspective. In this way, the technology and systems being developed could be adapted properly towards the non-human animal user. This is in a similar way that Human Computer Interaction (HCI) has done with human-being users.

Growth of Animal-Computer Interaction Research

The scientific aims, methodological approach and ethical principles of ACI were defined in Mancini's manifesto [9] in 2011 following on from the notable work of Resner [16]. Since then the ACI community has been growing and sparking the interest of more researchers diversifying into different focus areas such as zoo, pet and farming ACI, among others. This field has culminated into the first and second International Conferences on ACI, in 2014 and 2015 respectively, and an upcoming Third International Conference during 2016. There have also been several ACI workshops within computing and animal behavioral conferences, which has helped to spread the importance of this field to researchers of other areas. This diversifying is essential as when undertaking research within ACI there is need to provide knowledge from many different areas, ranging from computer science to animal behavior as well as ensuring all aspects of animal welfare.

There are many subfields within ACI holding notable works. A major area of ACI is to support animal welfare [15, 19], for example through devices such as automated kennels [11] or playful environments [13] in order to avoid stress or isolation problems. Similarly to Computer Science, the advancements in tracking technology have also been involved in ACI: from automated facial recognition for dogs [6] to body posture recognition in cats [14]. Game studies field has also approached animal play with cats [20], pigs [1], dogs [2], chickens [8], orangutans [21] and elephants [5]. Humans have also benefitted from ACI through systems which allow animals to assist them, usually through service dogs. We can also mention the development of interactive devices for aiding the detection of cancer with detection dogs [10], devices that support diabetes alert dogs when they need to alert their owner, or even emergency services [17]. Wearable technology has also been developed to facilitate remote communication between humans and dogs with occupations. These technologies include search and rescue dogs, allowing dogs to communicate using gestures which are detected by the wearable device [3]. More recently technology has been developed for zoo Animal-Computer Interaction, with a SIG (Special Interest Group) meeting being held at the 34th International Conference on Computer Human

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Interaction (CHI'16) to discuss on how to provide suitable technological enrichments for these animals [4].

Animal-Computer Interaction and Measuring Behavior

Due to the nature of the field and the uniqueness of its end-users, ACI researchers have to face several difficulties when designing for and with animals. One key difficulty is designing for non-verbal users, which HCI also faces. It is in this way that ACI and HCI share similar methods at gathering requirements and creating interactions. These methods, however, must be adapted towards the end user [7]. As animals mainly express themselves through non-spoken language, understanding and measuring their behavior is fundamental for the development of ACI research and methodologies. This assessment of behavior can be supported through other requirements gathering techniques such as biometric feedback, chemical analysis and owner evaluations. However, as the Internet of Things (IoT) increases into the everyday habits that we share with animals, systems that are able to automatically recognize, and more importantly react, to animal behavior are being built.

Collaboration of Animal Computer Interaction and Measuring Behavior

If we want to develop suitable technology for animals following a user-centered approach, our target users should be involved in all the stages of the design and development process. Hence, analyzing and evaluating their behaviors and responses towards the technology allows animals to be involved in the design process. Using this method, the two fields of animal behavior and computer science collide to create informative ACI. In this regards whilst ACI takes from animal behavior, ACI can also help inform animal behaviorists on working with technological systems and animals. In this way discussions will be had upon the methodological, theoretical, theme-based and focused-based approaches taken by the authors. The two fields through their exploration of how to measure animals' behavior can strengthen together. Whilst systems are being built that are able to measure animals' behavior it is only through correct interpretation that real meaning can be drawn from such interactions and thus analyzed by a system. From the above discussion it is clear that the interactions with technology so far remains interdisciplinary and needs input from a range of different fields (such as humanities, games studies, design research, animal sciences, social sciences to name a few). Through this symposium it is hoped to explore the relationship between ACI and animal-behavior. A possible research agenda for collaboration during and beyond this Symposium could be the following:

- First, a systematic review of existing literature in the field of measuring behavior regarding technological methods and approaches could be drawn with the help of experts in this field.
- Second, an evaluation of existing methods and work within ACI for measuring behavior could be assessed by experts in animal behavior, outlining potential pitfalls as well as strong points.
- Third, researchers in both fields could collaborate together to come up with areas in which measuring behavior by means of technological artifacts could be improved and/or provide benefits for the animal.
- Fourth, researchers could discuss which methodologies are used in their own area of expertise and whether or not these methodologies could be transferred from one field to another or redesigned following an animal-centred perspective.
- Fifth, discussion could be held to determine the extent to which technological artifacts could be a viable way of measuring behavior: how the technological artifacts might affect the interaction, how to interpret data into meaningful behavior, how autonomous could this behavior recognition process be, etcetera.

Symposium Animal-Computer Interaction and Measuring Behavior

The aims of this Symposium are twofold. Firstly, we want to introduce the topic Animal-Computer Interaction in the Measuring Behavior community with the assumption that there can be fruitful interaction between these two communities. Secondly, we aim at contributions that address methodological questions. How to conduct user-centred design in Animal-Computer Interaction or in computer-mediated human-animal interaction? Which methodologies from HCI can be adapted to ACI? Clearly, in this emerging field of research case studies can help to give answers to these questions as well.

The main topics that have been mentioned above also appear in the papers that have been accepted for this symposium. The papers and presentations address participatory and animal-centred design of playful and social interfaces for animals, the use of wearables (sensors) to monitor animals, technology that allows natural interactions, investigation animal preferences, trying to understand how animals experience and understand audio-visual stimuli, validating results of experiments, developing social relationships, et cetera. We present short summaries of the papers accepted for this workshop, emphasizing how they contribute to these issues.

Participatory design of interfaces to control playful applications is reported in “Playful UX for Elephants” by Fiona French and collaborators. Their research aims at implementing interactive toys and adaptive systems for captive elephants. Designed interfaces allow for shower control, audio control and musical instrument control. Issues that are investigated are interest, motivation and playfulness, and usability. This latter involves ‘natural’ control of applications using natural trunk movements and trying to discover an elephant’s preferences, for example the kind of haptic feedback to the trunk. In general, what kinds of systems have elephant appeal? Detecting trunk proximity (in order to allow trunk control of buttons) is one of the many technical challenges.

Explicit ethical guidelines for dog-computer interaction are discussed in the paper “The Ethics of How to Work with Dogs in Animal Computer Interaction” by Ilyena Hirskyj-Douglas and Janet Read. The guidelines follow from various case studies on Animal-Computer Interaction and animal centered and participatory design approaches, often done with dogs, including studies by the authors. The guidelines address issues such as consent, safeness and comfort, and avoidance of stress. Underlying the guidelines are insights and changing attitudes to cognitive and emotional capabilities of dogs, and comparisons with views held by HCI researchers on experiments with children. Within this paper, it is mentioned that a dog can never fully consent to take part within a study. One question that leaps out is whether a non-harm approach is always possible.

Naohisa Ohta and his co-authors discuss their ideas and research plans for investigating whether and how animals can collaborate with artificial presences or with humans through such presences in the contribution “Animal-Human Digital Interface: Can Animals Collaborate with Artificial Presences?”. It is argued that a multisensory approach with active agent functions is needed. More concrete research plans are mentioned. One of them is about the effect of including depth perception in artificial visual stimuli. In addition, there are plans to investigate whether social cues provided by robots can be understood and can help to develop social relationships between robots and animals. This will become an important issue when social and assistive robots become part of people’s and animals’ everyday life.

In the contribution “Animal-Computer Interaction (ACI): An analysis, a perspective, and guidelines” Egon van den Broek describes his experiences while reading the current ACI literature and his viewpoints on issues that should play a role in discussions on the methodological foundations for ACI research. Ideas follow from robot researcher Rodney A. Brooks’ views on the role of interaction and context in the development of artificial intelligence, and previous research on interactions between humans and artificial animal agents. Apart from this historical reflection and a subsequent emphasis on context-aware computing, van den Broek surveys approaches to validation of experimental results and the ‘triangulation’ techniques for validation that have emerged in the social sciences.

The paper “Towards a Wearer-Centred Framework for Animal Biotelemetry” by Patrizia Paci and co-authors focuses on how to design and use sensor technology in an animal-centred way. Body-attached sensors can track an animal’s activities (movements, body postures, absolute and relative positions) and provide physiological information. However, they can also interfere with the phenomena being studied. Sensors should conform to the wearer’s naturalistic setting and behavior. Otherwise there can be negative effects on the validity of the recorded data or the animal’s welfare. The authors have developed an animal-centred framework for designing wearable sensors with sensory imperceptibility, physical unobtrusiveness and cognitive acceptability as main parameters. In addition, it takes into account that such sensors can disturb natural interactions with other members of the same social group.

There are various ways to observe animal behavior. Indeed, we can use body-attached sensors as discussed in the Paci paper mentioned above. In "Detecting Animals' Body Postures Using Depth-Based Tracking Systems" Patricia Pons and co-authors refrain from accelerometers and gyroscopes attached to a dog's or cat's collar and instead use computer vision techniques (depth-based tracking using Microsoft Kinect) to distinguish locations and body postures (sitting, walking, jumping, et cetera) of a cat. It is preliminary work, presently with good results on cats, and the authors expect that more postures and behavior can be detected using machine learning algorithms and that the work can be extended to other animal species.

Conclusions

In this introductory paper to the Measuring Behavior Symposium Animal-Computer Interaction we introduced the main topics of research on ACI: anticipatory and animal-centred design of playful and social interfaces for animals, the use of wearables (sensors) to monitor animals, technology that allows natural interactions, investigation of animal preferences, trying to understand how animals experience and understand audio-visual stimuli, validating results of experiments, and understanding the development of social relationships. We think it is useful to introduce these issues in the Measuring Behavior community, find more interest and look for cooperation. It is expected that this Symposium will allow for fruitful collaboration between researchers from both areas as well as to provide insightful ideas for both the presented articles as well as for the future of the ACI and the Measuring Behavior communities.

References

1. Alfrink, K., Peer, I. van, Lagerweij, H., Driessen, C., Bracke, M., Copier, M. 2012. Pig Chase. Playing with Pigs project. www.playingwithpigs.nl. Accessed 15 March 2016.
2. Baskin, S., Zamansky, A. (2015). The Player is Chewing the Tablet! In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, 463–468.
3. Bozkurt, A., Roberts, D.L., Sherman, B.L., Brugarolas, R., Mealin, S., Majikes, J., et al. (2014). Toward Cyber-Enhanced Working Dogs for Search and Rescue. *IEEE Intelligent Systems* **29**(6), 32–39.
4. Carter, M., Webber, S., Sherwen, S. (2015). Naturalism and ACI: Augmenting Zoo Enclosures with Digital Technology. In: *2nd International Conference on Animal Computer Interaction, Proceedings of the 2015 Workshops on Advances in Computer Entertainment Conference - ACE '15 Workshops.*
5. French, F., Mancini, C., Sharp, H. (2015). Designing Interactive Toys for Elephants. In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, 523–528.
6. Hirskyj-Douglas, I., Luo, H., Read, J.C. (2014). Is My Dog Watching TV? In: *NordiCHI'14 - Workshop on Animal-Computer Interaction (ACI): Pushing Boundaries beyond "Human".*
7. Hirskyj-Douglas, I., Read, J.C. (2015). Doggy Ladder of Participation (DLOP). In: *Proceedings of the British Human Computer Interaction Conference - Animal Computer Interaction Workshop*, 1–4.
8. Lee, S.P., Cheok, A.D., James, T.K.S., Debra, G.P.L., Jie, C.W., Chuang, W., et al. (2006). A mobile pet wearable computer and mixed reality system for human–poultry interaction through the internet. *Personal and Ubiquitous Computing* **10**(5), 301–317.
9. Mancini, C. (2011). Animal-computer interaction: a manifesto. *Interactions* **18**(4), 69–73.
10. Mancini, C., Harris, R., Aengenheister, B., Guest, C. (2015). Re-Centering Multispecies Practices: A Canine Interface for Cancer Detection Dogs. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 2673–2682.
11. Mancini, C., Linden, J. Van Der. (2014). UbiComp for animal welfare: envisioning smart environments for kenneled dogs. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing,* 117–128.
12. Matsuzawa, T. (2003). The Ai project: historical and ecological contexts. *Animal Cognition* **6**(4), 199–211.

13. Pons, P., Jaen, J., Catala, A. 2015. Envisioning Future Playful Interactive Environments for Animals. In: Nijholt A, editor. *More Playful User Interfaces,* Springer, 121–150.
14. Pons, P., Jaen, J., Catala, A. (2015). Developing a depth-based tracking system for interactive playful environments with animals. In: *2nd International Conference on Animal Computer Interaction, Proceedings of the 2015 Workshops on Advances in Computer Entertainment Conference - ACE '15 Workshops.*
15. Rault, J., Webber, S., Carter, M. (2015). Cross - Disciplinary Perspectives on Animal Welfare Science and Animal - Computer Interaction. In: *2nd International Conference on Animal Computer Interaction, Proceedings of the 2015 Workshops on Advances in Computer Entertainment Conference - ACE '15 Workshops.*
16. Resner, B. 2001. Rover@Home: Computer mediated remote interaction between humans and dogs. M. Sc. Thesis, Massachusetts Institute of Technology.
17. Robinson, C., Mancini, C., Linden, J. Van Der, Guest, C., Harris, R. (2014). Canine-Centered Interface Design: Supporting the Work of Diabetes Alert Dogs. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3757–3766.
18. Rumbaugh, D. M. (1977). Language learning by a chimpanzee: the Lana Project. New York: Academic Press.
19. Väätäjä, H., Pesonen, E. (2013). Ethical issues and guidelines when conducting HCI studies with animals. In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 2159–2168.
20. Westerlaken, M., Gualeni, S. (2014). Grounded Zoomorphism: An Evaluation Methodology for ACI Design. In: *Proceedings of the 2014 Workshops on Advances in Computer Entertainment Conference - ACE '14 Workshops.* ACM, Article 5, 6 pages.
21. Wirman, H. (2014). Games for/with Strangers: Captive Orangutan (Pongo Pygmaeus) Touch Screen Play. In: *Antennae: The Journal of Nature in Visual Culture* **30**, 103-113.

Transfer Learning for Rodent Behavior Recognition

M. Lorbach[1,2], R. Poppe[1], E.A. van Dam[2], L.P.J.J. Noldus[2] and R.C. Veltkamp[1]

[1]Utrecht University, Utrecht, The Netherlands.
m.t.lorbach@uu.nl
[2]Noldus Information Technology, Wageningen, The Netherlands

Most automated behavior recognition systems are trained and tested on a single dataset, which limits their application to comparable datasets. While training a new system for a new dataset is possible, it involves laborious annotation efforts. We propose to reduce the annotation effort by reusing the knowledge obtained from previous datasets and adapting the recognition system to the novel data. To this end, we investigate the use of transfer learning in the context of rodent behavior recognition. Specifically, we look at two transfer learning methods and examine the implications of their respective assumptions on synthetic data. We further illustrate their performance in transferring a rat action classifier to a mouse action classifier. The performance results in the transfer task are promising. The classification accuracy improves substantially with only very few labeled examples from the novel dataset.

Introduction

The automated recognition of rodent behavior plays an important role in studies of neurological disorders such as Huntington's disease. With the introduction of automated recognition systems, the results of such studies have become more accurate and better reproducible across laboratories [1].

Automated behavior recognition systems typically track the animals in videos and classify their behavior into several action categories. This classification involves an action model that is typically learned from annotated training sequences. Often these training sequences originate from a single dataset of one specific study [2]–[6]. As a consequence, a model that is applied to a dataset that is different from the training dataset may cause unwanted bias in the results [7].

In practice, every dataset is different. The data are affected by a variety of factors that occur in animal studies. For instance, laboratories use different video acquisition setups and cages. The animals behave differently depending on their species, age, gender and individual traits, but also on their treatment and potential disease progress. Each of these factors has an effect on the data distributions and therefore potentially on the accuracy of the recognition system.

To ensure high recognition accuracy on a novel dataset, we could train a new action model tailored to that dataset [8]. However, the training comes at the cost of manual annotation efforts. Furthermore, a new action model would not take previous models into account and would therefore neglect valuable knowledge. Instead of learning a new action model from scratch, we propose to adapt the existing model to the new dataset by adopting a transfer learning approach.

The key idea of transfer learning is to build upon previously obtained knowledge and combine it with a small amount of additional training data to obtain a model that performs optimally with respect to a novel dataset. The additional training data may contain annotated but also unannotated instances from the novel dataset depending on the method. In order to be able to balance previous and new knowledge, transfer learning methods make varying assumptions about which parts of the model do and do not change across datasets.

In this paper, we take the first steps towards a behavior recognition system that combines previously learned knowledge with the ability to adapt to new datasets. We investigate the strengths and limitations of two transfer learning methods with fundamentally different assumptions in the context of rodent behavior recognition. We

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

examine the implications of their respective assumptions on synthetic data and show their performance in a task transferring an action model trained on rats to a new action model for mice.

Related Work on Rodent Behavior Recognition

The automated recognition of rodent behaviors typically involves classifying video frames into behavior categories based on features extracted from the video. The features may contain the trajectories of the moving animals as well as shape, pose and distance information. The system's performance is eventually evaluated on a test set by comparing the automated classification to the annotations of one or multiple humans.

A large number of recognition systems for animal behavior are trained and evaluated on subsets of one dataset [2]–[6]. Using different strategies, the dataset is split into training and evaluation sets. To increase the statistical power of the evaluation, multiple splits can be generated randomly and the accuracy values are then averaged [9]. Despite their statistical power, the scope of reported performance is limited to one specific dataset and we are uncertain about the validity on other datasets.

The performance can be considerably lower on other datasets. An automated behavior recognition method for rats, for instance, has been reported to show a decrease in average precision by 17% if the system is tested on animals it has not encountered during training compared to when it was trained and tested on the same animal [10]. Along this line of thought, another study has recently shown that individual animals have distinct modes in their velocity and distance values [11]. Such individual traits potentially affect the automatic classification of, for example, walking and running.

In addition, the experiment environment can influence the performance. For instance, illumination and viewpoint variations may affect the animal tracking whereas the dimensions of the cage scales distance and velocity features. Although some features such as distances can be normalized [12], factors such as individual traits of the animals do not correspond to single features and are therefore difficult to compensate for manually. A possible countermeasure is to include the expected variations of the environment and the animals in the dataset [9], [12]. Although such variations can improve the generalization properties of a classifier, they are limited to the included variations. Clearly, it is infeasible to create an annotated dataset that includes the infinite number of real-world variations and then test on it efficiently. Moreover, data with high variance pose classification challenges that can only be countered by complex classifiers and even more data.

In this work we investigate the task of adapting a classifier that has been trained on one dataset to a novel dataset. In the field of machine learning such a task is considered a transfer learning problem.

Transfer Learning

Transfer learning addresses classification and regression scenarios in which the training and the test data are sampled from different data distributions [13]. Such scenarios occur frequently in real-world applications. For instance, for medical diagnosis a disease model may be learned from a group of affected patients and an equally sized group of healthy volunteers. If such a model is later applied to predict the disease in a much larger and more general population, it is likely to cause biased results. In this particular case, the model suffers from a sample selection bias.

Sample Selection Bias

Sample selection bias can degrade the accuracy of a model considerably [14]. We distinguish between two main sources of how a sample may become biased. First, as in the example of the disease model, the training set contains relatively more positive examples than the true population. The model is biased towards the positive diagnosis. In this case, the sample selection is dependent on the label (positive or negative).

Second, the sample selection can also depend on the feature values. Consider the task of classifying fish species based on their length. If we take the training examples from a pool of mainly juvenile fish, we obtain a distorted image of the true distribution of length values. Again, classifications we make on basis of that model are biased.

Once we are aware of a sample selection bias in our training data, we can attempt to correct for it using a transfer learning approach. In transfer learning we consider data coming from different domains. While labeled training data is obtained from the source domain, the test data comes from the target domain. The transfer learning problem is then formulated as the task of finding a classifier that performs optimally in the target domain.

Several methods address sample selection bias. The majority assumes that we do not have any labeled instances from the target domain to guide the knowledge transfer across the domains. Although we rely on training the model on source data, we can influence the training using the *unlabeled* target data. For instance, we can concentrate our learning effort on source samples that are surrounded by many target samples in the feature space and neglect the deviating samples. Assuming that the labels between the selected samples are the same, we yield a suitable model of the target domain. We need to ensure, however, that all target samples are close to at least some source samples so that all target samples are eventually reflected in the model.

The concept of identifying the most important source samples is the key idea behind Kernel Mean Matching (KMM) [15], [16]. KMM distributes weights to the source samples according to their importance by minimizing the difference between the means of the target data and the weighted source data. Since the weighted source data match the target data better, a classifier trained on the weighted data performs better in the target domain.

A similar approach is taken by Transfer Component Analysis (TCA) [17]. Instead of weighting source samples, TCA finds a feature mapping that minimizes the difference between source and target distributions in the mapped feature space.

Regardless of the source of the bias, a sample selection bias influences the marginal probability distributions. As a consequence the distribution over samples X differ between domains, i.e., $P_{Source}(X) \neq P_{Target}(X)$. The key assumption that enables us to compensate for these differences is that the conditional probability distributions are not affected by the bias. That is, the probability of the occurrence of a label y given a sample x is the same irrespective of the domain: $P_{Source}(y|x) = P_{Target}(y|x)$.

General Dataset Shift

In the case that the assumption about the conditional distribution does not hold, we are facing a more general dataset shift problem. To solve it, we typically need to make other assumptions about the domain differences [18] or introduce information about the target domain, for example, by providing labeled target instances [19].

The availability of labeled target instances enables us to estimate and optimize the performance of a classifier in the target domain. The transfer AdaBoost (trAdaBoost) [20] method repeatedly trains a classifier from the source data and predicts the labels of the known target instances. In each iteration it removes source instances that did not contain valuable information for classifying the target samples. After a number of repetitions the classifier is left with only the most important source samples and hence performs optimally in the target domain.

As opposed to manipulating training instances and feature representations to change the data model, we are also able to directly manipulate the classifier and its parameters. The practical advantage is that the original source data does not need to be available for the transfer [21]. The disadvantage is that introducing new labeled examples involves new challenges such as finding the right balance between the established source model and the new target model [22].

An example of a method that manipulates an existing classifier is the adaptive Support Vector Machine (aSVM) [23]. The prerequisite for aSVM is a classifier that has been trained on the source data. aSVM then changes the decision function of that classifier such that the classification error regarding the target instances is minimized.

At the same time, the modification of the decision function is kept as small as possible. The balance between classification error and degree of modification is controlled via a parameter that is determined manually.

For the purpose of investigating transfer learning in the context of rodent behavior recognition, we concentrate our analysis on two methods that take two fundamentally different approaches: KMM and aSVM. First, we examine their performance on synthetic data with a particular focus on violating their key assumptions. Second,

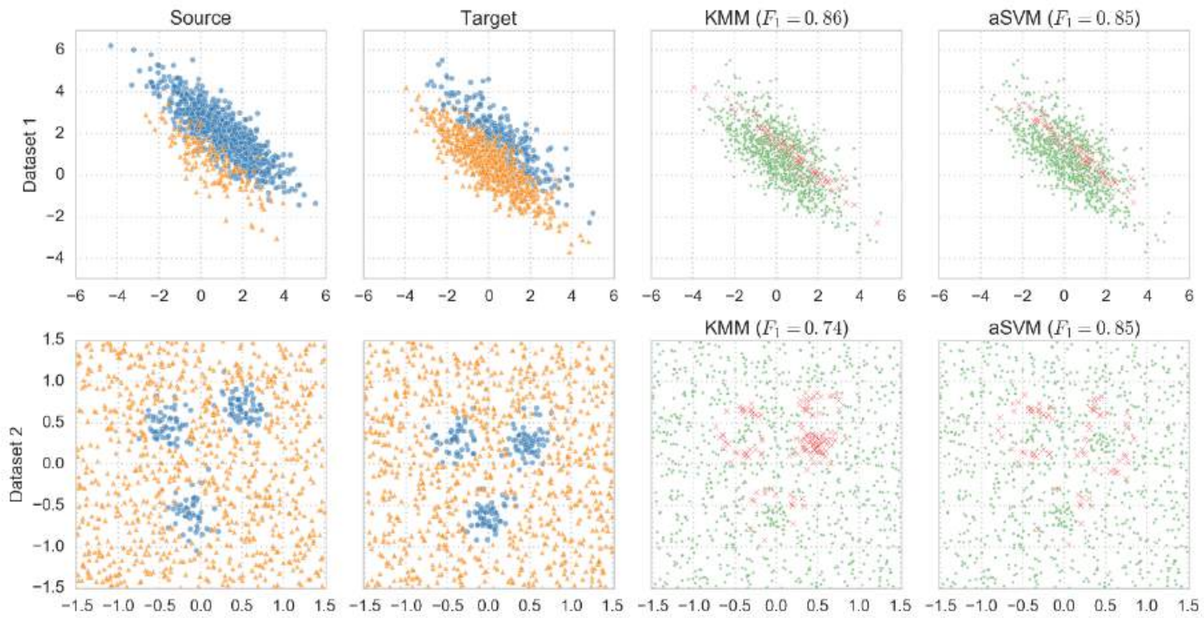


Figure 6: Left: Two synthetic datasets with two classes each are split into a source and a target set. Right: the correct (green dots) and incorrect (red crosses) predictions of KMM and aSVM (5% labeled target instances) on the target data.

we apply them to a simplified rat to mouse behavior transfer problem and discuss their performance.

Experiments on Synthetic Data

We apply the two transfer learning methods, KMM and aSVM, to two synthetic datasets. The datasets are illustrated in **Figure 6**. Dataset 1 contains two classes each represented by a bivariate Normal distribution (class 1: $\mu = [1 \quad 2], \sigma = \begin{bmatrix} 1.7 & -1.2 \\ -1.7 & 1.7 \end{bmatrix}$, class 2: $\mu = [0.5 \quad 0.6], \sigma = \begin{bmatrix} 1.7 & -1.2 \\ -1.7 & 1.7 \end{bmatrix}$). The source dataset is sampled from the true distribution under a sampling bias as proposed in related work [14]. Samples of class 1 are selected with a probability of 0.85 ($N_1 = 850$) while samples of class 2 with a probability of 0.15 ($N_2 = 150$). The target dataset consists of $N_1 = 300$ and $N_2 = 700$ samples and the class means are shifted by $\Delta\mu = [-0.3 \quad -0.2]$. Dataset 1 therefore includes a sample selection bias and a small shift of the conditional probability distribution.

Dataset 2 has been used in related work to illustrate the aSVM method [23] and consists of a positive and a negative class. The positive class is represented by a mixture of three bivariate Normal distributions while the negative class is distributed uniformly outside the positive class. In the source domain, the positive class is determined by the parameters $\mu_1 = [-0.4 \quad 0.5], \mu_2 = [0.5 \quad 0.7], \mu_3 = [-0.1 \quad -0.6], \sigma = 0.02$; in the target domain by $\mu_1 = [-0.4 \quad 0.3], \mu_2 = [0.5 \quad 0.3], \mu_3 = [0 \quad -0.65], \sigma = 0.02$. Both source and target sets comprise 166 positive samples and 834 negative samples. Dataset 2 does not suffer from a sample selection bias but has a considerable shift in the conditional probability distributions.

Before training the datasets are scaled to zero-mean and unit-variance based on the source data only. KMM is applied to the source and the entire, unlabeled target data to obtain the sample weights for the source samples. A SVM is then trained on the weighted source samples with parameters for dataset 1: linear kernel, C=1; and for

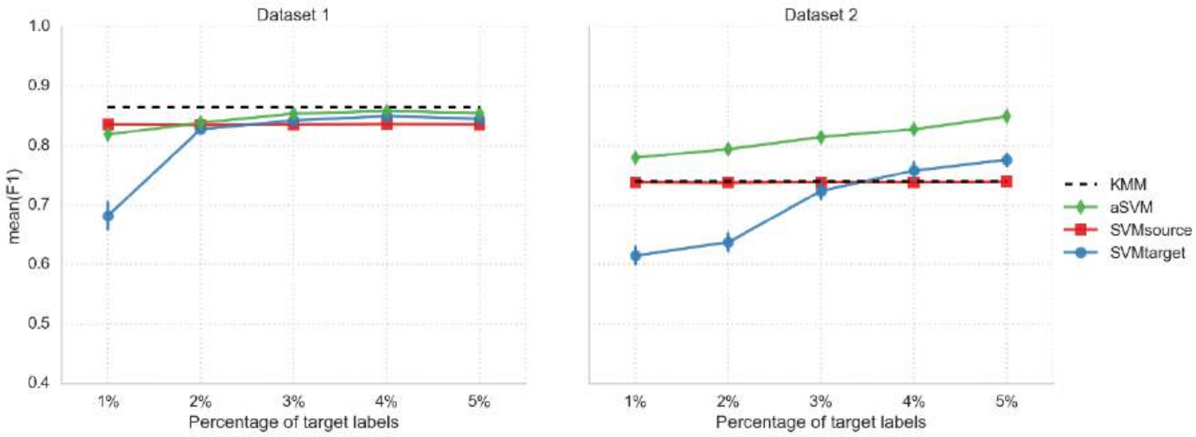


Figure 7: Accuracy on synthetic datasets with varying amount of labeled target instances. Error bars show the standard error of the mean. Number of target samples (100%): N=1000.

dataset 2: radial basis function (RBF) kernel, C=1, $\gamma = 5$.

For aSVM, we first train an SVM on the source data with the same parameters as before. Then, aSVM is applied to adapt the trained classifier given a varying amount of labeled target instances. We provide 1%, 2%, 3%, 4% and 5% of the target set as labeled instances and subsequently test on the remaining samples. We perform 30 repetitions using randomized samples and average the results among the repetitions.

We evaluate the performance of the two methods in terms of the F1 score averaged across classes. For comparison, we include the performance of two baseline classifiers that do perform any knowledge transfer. First, an SVM is trained only on source data (SVMsource) and second, an SVM is trained only on the available target data (SVMtarget). All transfer learning methods should outperform both baseline classifiers.

Results

The averaged accuracies of the methods are shown in **Figure 7**. On dataset 1, KMM outperforms the other methods slightly. The accuracy of aSVM is comparable to SVMsource with only 1% of labeled target instances. With more data, aSVM converges to a similar decision boundary as KMM with comparable classification errors (**Figure 6**). If the SVM is trained on the labeled target data alone, we need at least 2% of the data to reach a performance comparable to the other methods.

On dataset 2, KMM does not improve the prediction accuracy over the SVM that is trained on the source samples. The SVMtarget outperforms SVMsource if it has access to more than approximately 3.5% of the target data. aSVM achieves the highest accuracy of the compared methods. The accuracy of aSVM increases as more target instances become available. The increase is particularly strong on this dataset. The source model can be improved even with very few target instances.

Discussion

KMM is able to compensate for the sample selection bias and the small shift in the features in dataset 1. Due to the weighing of source samples, the resulting classification model is a good fit for the target data. In contrast, on dataset 2, the assumption that the conditional distributions remain the same across domains is violated and consequently KMM misclassifies more samples. Eventually, the lack of labeled target instances prevents KMM from improving the original source model.

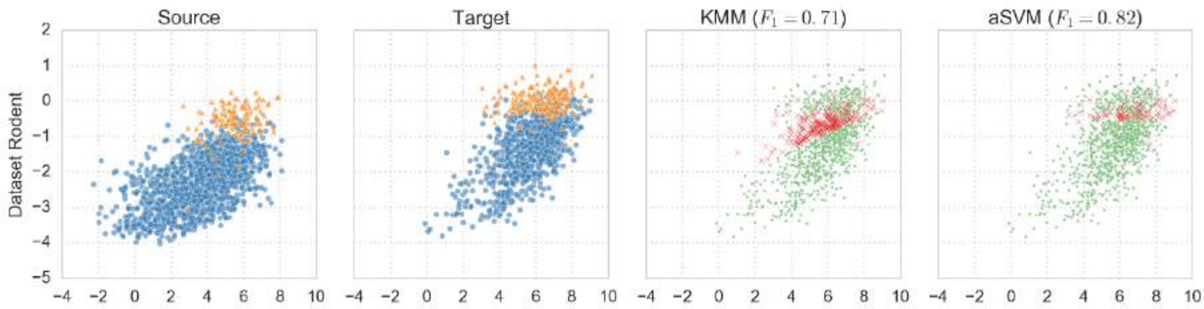


Figure 8: Left: Rodent behavior dataset with two classes *sniff* (blue circles) and *walk* (orange triangles). Right: the correct (green dots) and incorrect (red crosses) predictions of KMM and aSVM (1% labeled target instances) on the target data.

aSVM is able to use the available target labels to improve the source model. Moreover, aSVM outperforms SVMtarget which shows that the combination of knowledge from a previously trained model and knowledge from new instances is superior to learning a new model from scratch.

The experiments on synthetic data show that the KMM is too sensitive to violating its key assumptions. aSVM appears more robust to changes in the conditional distributions and may therefore be better suited for a real-world application such as rodent behavior recognition.

Experiments on Rodent Behavior Data

We now apply the two transfer learning methods to a simplified dataset of rodent behaviors. The learning task consists of transferring an action model for rats (source domain) to mice (target domain). Rats and mice perform similar actions but differ for example in their size and velocity. These differences affect the marginal as well as the conditional probability distributions.

The simplified dataset comprises a subset of a larger rodent behavior dataset. Every sample in the dataset corresponds to a video frame from which a number of features are extracted. For the purpose of this article we select two actions *sniff* and *walk*, and two features *Velocity* and *Optical Flow Energy*. The chosen subset reflects a scenario in which the shift in the conditional distribution causes misclassifications in practice. The source set contains $N_{sniff} = 1261$ (91%) and $N_{walk} = 127$ (9%) samples of sniffing and walking actions of one rat, respectively. The target set contains $N_{sniff} = 820$ (81%) and $N_{walk} = 194$ (19%) samples of one mouse.

The rodent dataset poses several classification challenges. The occurrence of the classes is highly unbalanced with a small selection bias towards *sniff* in the source domain. Furthermore, the conditional distributions differ between the domains as visualized in **Figure 8**.

We apply both KMM and aSVM analogous to the synthetic datasets. Source and target SVMs are parameterized with a linear kernel (C=1). The training is performed with a varying amount of labeled target instances and is repeated 30 times with randomized samples. The performance is evaluated in terms of the F1 score averaged over classes.

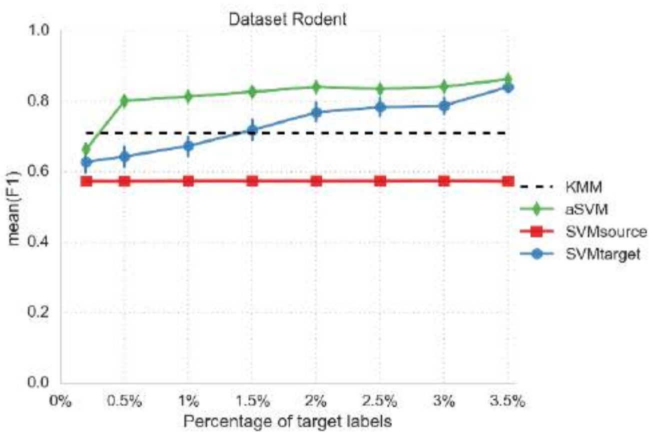


Figure 9: Performance on real rodent behavior data with varying amount of labeled target instances. Error bars show the standard error of the mean. Number of target samples (100%): N=1014.

Results

The accuracy of aSVM increases with the number of available labeled target instances (**Figure 9**). Similar to the synthetic dataset 2, only five labeled samples are sufficient to reach a substantial increase in performance over both SVMsource and SVMtarget. With approximately 3.5% of the target instances being labeled, SVMtarget has sufficient information to perform almost as well as aSVM.

KMM improves the accuracy of the original source model by approximately 0.14 but is outperformed by aSVM with only 0.5% labeled target instances.

In **Figure 8** we count more misclassifications by KMM compared to aSVM. aSVM manages to find a decision boundary that matches the classes in the target domain well. A few classification errors remain along the decision boundary.

Discussion

The main challenge of the rodent dataset is similar to the challenge in synthetic dataset 2, namely the shifted conditional distribution. Consequently, aSVM is, as in the synthetic case, able to deal with that challenge better than KMM. aSVM therefore presents a viable option for transferring classifiers to another domain even if only very few labeled target instances are available.

In all analyzed datasets, the transfer of knowledge from the source to the target domain is achieved best by aSVM. Although it cannot quite match the accuracy of KMM on the synthetic dataset 1, aSVM is more efficient to compute. While KMM evaluates the pairwise distances between all source and target samples, the calculations for aSVM are limited to minimizing the expected loss over the small amount of labeled target samples. The efficiency of aSVM will be an advantage if knowledge is to be transferred among larger datasets.

In the performed experiments, we posed relatively simple classification problems. The classifiers' performances saturate with very few examples so that the differences between the models vanish quickly. Moreover, we have only looked at binary classification tasks in a two-dimensional feature space. In rodent behavior recognition we need to distinguish between 8 or more classes in much higher dimensions. Further investigations are needed in order to determine the limits of aSVM.

A practical aspect that requires our attention is a suitable strategy for sampling labeled target instances. In this work, we sampled uniformly from the available target data. For rodent behavior recognition, annotating single, random video frames may not be the best strategy as it is important to capture the variance within the performance of an action as well as across such performances. Therefore, annotating multiple, continuous segments of different behaviors may be a better approach.

A related question is whether we can make the learning more efficient by avoiding random sampling and rather focusing on the most informative segments first. If this selection process could be automated then the user could be queried to provide annotations for the informative segments. The adaptation of the recognition system to a new dataset could then be carried out interactively with the user.

Conclusion

We have investigated transfer learning in the context of rodent behavior recognition. With transfer learning we aim at exploiting previously obtained knowledge from other datasets in order to improve the recognition in a novel dataset while reducing the annotation effort. In this article, we have evaluated two transfer learning methods and examined the implications of their respective assumptions on synthetic data. We have further illustrated their performance in transferring a rat action classifier to a mouse action classifier.

The results of aSVM on the rodent behavior transfer task are promising. The fact that very few labeled examples from the novel dataset are sufficient to substantially improve the classification accuracy, encourage further investigations using aSVM. In the future, we will evaluate its performance on more complex data with more classes and in higher dimensional feature spaces.

In its current implementation, aSVM can only handle binary classification problems. For its application to a wider range of rodent behavior recognition problems, the method has to be extended to multi-class problems. We will concentrate our efforts on extending the approach to more realistic classification tasks.

With our investigations, we have made the first step in the development of a rodent behavior recognition system that is adaptive to novel datasets under reduced annotation efforts. Further steps will be taken to enhance the applicability to more complex problems and to include the interaction with the user for higher efficiency.

Acknowledgement

This work was funded by the European Commission under the 7th Framework Programme (FP7) as part of the Marie Curie Initial Training Network ``PhenoRat'' (GA no. 317259).

References

1. F. A. Desland, A. Afzal, Z. Warraich, and J. Mocco, “Manual versus Automated Rodent Behavioral Assessment: Comparing Efficacy and Ease of Bederson and Garcia Neurological Deficit Scores to an Open Field Video-Tracking System,” *J. Cent. Nerv. Syst. Dis.*, vol. 6, pp. 7–14, 2014.
2. H. Dankert, L. Wang, E. D. Hoopfer, D. J. Anderson, and P. Perona, “Automated monitoring and analysis of social behavior in Drosophila,” *Nat. Methods*, vol. 6, no. 4, pp. 297–303, 2009.
3. E. Eyjolfsdottir, S. Branson, X. P. Burgos-Artizzu, E. D. Hoopfer, J. Schor, D. J. Anderson, and P. Perona, “Detecting Social Actions of Fruit Flies,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, vol. 8690, pp. 772–787.
4. W. Hong, A. Kennedy, X. P. Burgos-Artizzu, M. Zelikowsky, S. G. Navonne, P. Perona, and D. J. Anderson, “Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning,” *Proc. Natl. Acad. Sci.*, vol. 112, no. 38, pp. E5351–E5360, 2015.
5. X. P. Burgos-Artizzu, P. Dollár, D. Lin, D. J. Anderson, and P. Perona, “Social behavior recognition in continuous video,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1322–1329.
6. L. Giancardo, D. Sona, H. Huang, S. Sannino, F. Managò, D. Scheggia, F. Papaleo, and V. Murino, “Automatic Visual Tracking and Social Behaviour Analysis with Multiple Mice,” *PLoS ONE*, vol. 8, no. 9, p. e74557, 2013.
7. J. M. Girard and J. F. Cohn, “A Primer on Observational Measurement,” *Assessment*, 2016.
8. M. Kabra, A. A. Robie, M. Rivera-Alba, S. Branson, and K. Branson, “JAABA: interactive machine learning for automatic annotation of animal behavior,” *Nat. Methods*, vol. 10, no. 1, pp. 64–67, 2012.
9. H. Jhuang, E. Garrote, X. Yu, V. Khilnani, T. Poggio, A. D. Steele, and T. Serre, “Automated home-cage behavioural phenotyping of mice,” *Nat. Commun.*, vol. 1, no. 6, pp. 1–9, 2010.
10. J. B. I. Rousseau, P. B. A. Van Lochem, W. H. Gispen, and B. M. Spruijt, “Classification of rat behavior with an image-processing method and a neural network,” *Behav. Res. Methods Instrum. Comput.*, vol. 32, no. 1, pp. 63–71, 2000.
11. S. M. Peters, I. J. Pinter, H. H. J. Pothuizen, R. C. de Heer, J. E. van der Harst, and B. M. Spruijt, “Novel approach to automatically classify rat social behavior using a video tracking system,” *J. Neurosci. Methods*, 2016.
12. E. A. van Dam, J. E. van der Harst, C. J. F. ter Braak, R. A. J. Tegelenbosch, B. M. Spruijt, and L. P. J. J. Noldus, “An automated system for the recognition of various specific rat behaviours,” *J. Neurosci. Methods*, vol. 218, no. 2, pp. 214–224, 2013.
13. S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.
14. B. Zadrozny, “Learning and Evaluating Classifiers under Sample Selection Bias,” in *International Conference on Machine Learning (ICML)*, 2004, pp. 903–910.
15. J. Huang, A. Smola, A. Gretton, K. Borgwardt, and B. Schölkopf, “Correcting Sample Selection Bias by

Unlabeled Data," in *Advances in Neural Information Processing Systems*, 2007, pp. 601–608.
16. M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. von Bünau, and M. Kawanabe, "Direct importance estimation for covariate shift adaptation," *Ann. Inst. Stat. Math.*, vol. 60, no. 4, pp. 699–746, 2008.
17. S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *Neural Netw.*, vol. 22, no. 2, pp. 199–210, 2011.
18. K. Zhang, B. Schölkopf, K. Muandet, and Z. Wang, "Domain Adaptation under Target and Conditional Shift," in *International Conference on Machine Learning (ICML)*, 2013, pp. 819–827.
19. X. Wang and J. Schneider, "Flexible transfer learning under support and model shift," in *Advances in Neural Information Processing Systems*, 2014, pp. 1898–1906.
20. W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for Transfer Learning," in *International Conference on Machine Learning (ICML)*, 2007, pp. 193–200.
21. I. Kuzborskij and F. Orabona, "Stability and Hypothesis Transfer Learning," in *International Conference on Machine Learning (ICML)*, 2013, pp. 942–950.
22. S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A Theory of Learning from Different Domains," *Mach Learn*, vol. 79, no. 1–2, pp. 151–175, 2010.
23. J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain Video Concept Detection Using Adaptive SVMs," in *International Conference on Multimedia (MM)*, 2007, pp. 188–197.

Validation of Novel Tools to Analyse Behavior in a Rat Model of Spinocerebellar Ataxia Type 17

Elisavet Ioanna Kyriakou[1,2,3], Jesus Arellano Spadaro[1], Alex van der Starre[1], Johanneke E van der Harst[1,2], Judith R Homberg[1] and Huu Phuc Nguyen[3]

[1]Radboud University, Nijmegen, The Netherlands
[2]Noldus Information Technology, Wageningen, The Netherlands
[3]Tuebingen University, Tuebingen, Germany

Introduction

Neurodegenerative disease is a broad term used to describe disorders which have, as a common characteristic, the progressive death of neurons and their inability to be reproduced or replaced. Depending on the area of the brain where the nerve cell death takes place, neurodegeneration can cause problems with movement or mental functioning. Movement disorders are known as "ataxia" and cognitive disorders are referred to as "dementia". In patients these two disorders are very often present at the same time [1]. Compared to movement disorders, cognitive aspects of dementia have been most extensively studied in rodents, resulting in a relatively limited understanding of movement disorders. In response to this limited understanding, an increasing number of animal models for movement disorders have been reported in literature during the past 10 years. However, proper interpretation of results is hindered by the lack of an objective and automated method to assess motor coordination and gait abnormalities [2]. The CatWalk system can provide an extensive number of parameters that allows automated assessment of gait and locomotion. The concept of CatWalk is based on the classical footprint analysis test in which the animal's paws were dipped into non-toxic paint or ink and then had to walk along a paper-covered corridor leaving a track of footprints. However, compared to the inked-paw version which can give mainly information about the static gait parameters, the CatWalk system can provide information about distribution of the animal's weight across its feet, the maximum surface area of a foot touching the ground, the intensity of a print, and most importantly, the temporal relations of the animal's paws such as the stand duration, the swing, the walking speed and so on [3], [4]. In addition high-throughput behavioural tests implemented in an automated home-cage environment are of great value towards the better understanding of genotype-phenotype relations. The light-spot test that uses a mild aversive stimulus in the home-cage (PhenoTyper) provides the benefit of circumventing human handling and monitoring the animal for a long period of time. [5]. Therefore, our aim is through these two tests to determine the disease's progression over time and identify the translational value of the behaviours identifies to the symptomatology of the patients suffering from SCA17 as described in literature[6]–[10].

Materials and Methods

Eight adult male SCA17 rats and eight wildtype Sprague Dawley rats at the age of 9 months old were used for the CatWalk experiment. For the home cage monitoring with the light spot task, 72 rats of 3,6, and 9 months old. Each age group consisted of 12 SCA17 and 12 wildtype Sprague Dawley rats. All animals were housed under standard conditions with 12:12 h reversed light:dark cycle (8:00AM-8:00PM red lights on; 8PM-8AM white lights on), temperature 21ºC (±2) housed in Makrolon type IV-S cages (Tecniplast, Italy) with 2 rats/cage. Food and water were available *ad* libitum.

Animals gait analysis was performed by using CatWalk XT 10.5 (Noldus Information Technology, Wageningen, the Netherlands). The apparatus consists of an enclosed corridor with a glass plate floor, and a goalbox at the end under which the home-cage of the animals can be put. The runway is illuminated from the ceiling with a red and green light. When the animals' paws have contact with the glass plate the light gets reflected and is captured by a high-speed video camera fixed 60 cm below the corridor. The width of the corridor was adjusted at approximately 8 cm to prevent the animals to turn back and interrupt their straight movement. The green

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

intensity threshold was set to.12 and the camera gain to 15.84. The animals were trained for three days both during their dark phase under red light conditions. To test and train the animals on the apparatus we used an alternative schedule. That is, the animals were tested in pairs with each animal having one run at a time, then their pair was put for one run, followed by the first animals again etc. This procedure was continues until at least 5 straight and with no interruption runs were acquired per animal. Also, the animals were motivated to traverse the corridor by using a food reward after entering their homecage underneath the goal box.

The light spot test was performed by using PhenoTyper 4500 and EthoVision XT 9 (Noldus Information Technology, the Netherlands). The home cage system consists of a cage with the size of 45x45cm equipped with a shelter for the animals to rest, a food hopper and a water spout. Video tracking is performed by an infrared video camera installed in the top unit attached on top of each cage separately. The light spot is a white led light of an intensity of 500 lux that shines for 3h after the start of the dark light (08:15AM) directly at the feeding area on day 6 of the home cage monitoring of the animals. The animals were habituated in the PhenoTyper for 3 days. Then the next 2 days the animals were monitored analysing their behaviour as a “natural behaviour” in their home environment. Next, on day 6 the light spot task takes place. Parameters such frequency and time spent in the shelter, moving, and not moving activity, distance moved, frequency and time spent in the feeder were analysed between the SCA17 and the wildtype animals to test for the behavioural response to this mild aversive stimulus within a habituated environment.

Results

Both static and dynamic parameters showed significant differences, with the SCA17 rats showing coordination disturbances. Although training affected some gait parameters, it did not obscure group differences when those were present.

Results analyzed from the transgenic and control group showed that the regularity of index was not affected (data not shown). On the contrary, there was a significant increased distance between the two forepaws ($F(1, 13) = 19.84$, $p = 0.001$; Fig. 1A) whereas the base of support of the hind paws was significantly decreased compared to the control group ($F(1, 13) = 9.8$, $p = 0.008$; Fig. 1A). In addition, maximal contact area was impaired in the SCA17 rats for the front ($F(1, 13) = 6.14$, $p = 0.028$) and hind paws ($F(1, 13) = 33.5$, $p < 0.0001$) proving a larger contact of the paws with the glass plate compared to the control. However, the print area remained unaffected. No differences were noted regarding other paw parameter such as the toe spread, print width and intensity

Concerning the temporal relations no difference in the average walking speed and the swing and stand duration was noted (data not shown). A significant decrease in the phase dispersion between the two diagonal paws indicates a disturbed inter-paw coordination as that the placement of the diagonal hind paws relative to the front paws were placed significantly earlier compared with the control group (diagonal pairs, RF(anchor) → LH(target)$F(1, 13) = 21.09$, $p = 0.001$; LF(anchor) → RH(target)$F(1, 13) = 6.56$, $p = 0.024$; Fig. 1E).

Despite the fact that stride length was significantly increased in both groups over the sessions (front paws, $F(2, 26) = 6.14$, $p = 0.007$; hind paws, $F(2, 26) = 8.18$, $p = 0.002$) no group difference was noted (Fig. 1B). In addition, the stand index was significantly decreased for both test groups over the three sessions from the first to the third day (front paws, $F(2, 26) = 9.05$, $p = 0.001$; hind paws, $F(2, 26) = 6.55$, $p = 0.005$; Fig. 1C). Concerning temporal relations of the paws, only the swing speed showed a session effect with the SCA17 rats showing an increase in the speed of the paw during no contact with the glass plate (front paws, $F(2, 26) = 10.5$, $p < 0.0001$; hind paws, $F(2, 26) = 6.4$, $p = 0.005$; Fig. 1D). Finally, regarding the fore paws’ maximal contact area there was only a session effect with the values manifesting a decrease (front paws, $F(2, 26) = 3436.3$, $p < 0.0001$; hind paws, $F(2, 26) = 5.67$, $p = 0.009$; data not shown). Potential effects of body weight on gait parameters were assessed and no significant correlation was found.

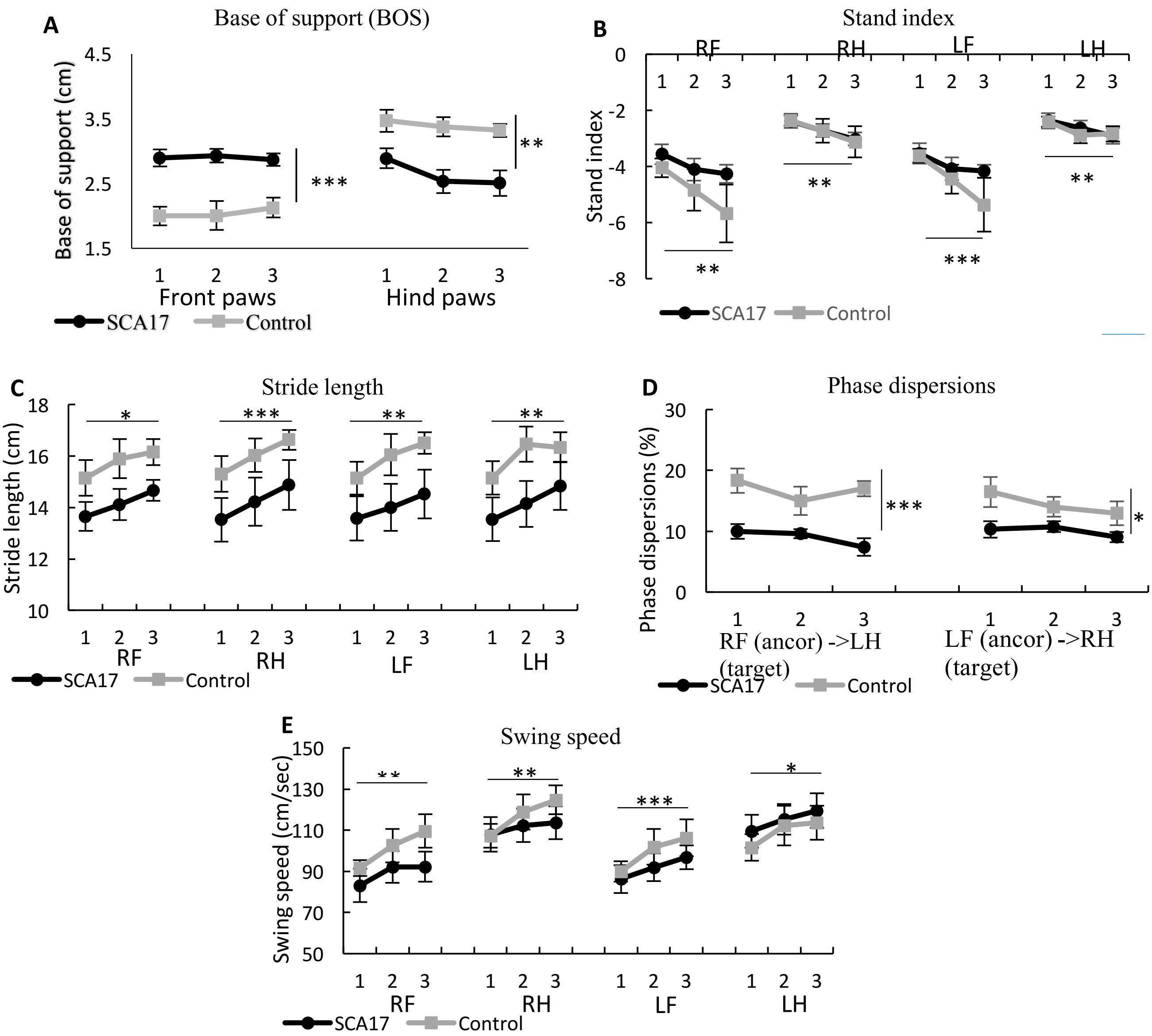


Figure 1: Motor phenotype of SCA17 rats. The CatWalk was performed at the age of 9 months when the disease has already progressed and the phenotype has been fully developed. Transgenic rats (black line) showed a significant higher base of support for the fore paws, in contrast with the distance between the hind paws that is significantly decreased (A). No significant group effect was detected for the stand index, stride length and the swing speed. All three parameters showed a significant time effect with the stand index (B) to exhibit a decrease over time, and the stride length (C) and swing speed (E) to show an increase throughout the three testing days in CatWalk. The phase dispersion was significantly affected (D), with the transgenic rats showing a faster placement of the hind paw on the glass plate in relation to the respective diagonal front paw as a sign of impaired gait coordination. All tests were performed with male rats. Gait parameters were assessed in three days, (n=6 controls; n=9 TBPQ64 rats). With each point representing the mean ±S.E.M. statistically significant differences in comparison with the control group. (two-way repeated measures ANOVA) *$p<0.05$; **$p<0.01$; ***$p<0.001$.

As SCA17 rats appeared to have impaired gait and locomotor skills, we subsequently evaluated the behavioural activity in the novel light spot test in search of an anxious response. We observed behaviour 24 hours before the light spot (baseline conditions) and on the day of the light spot test for the duration of 3 hours. Analysis of the total time spent in the feeder in normal conditions and the day of the light spot test, SCA17 rats on average spent more time in the feeder than their wildtype counterparts which can be observed at 6 and 9 months. To correct for real numbers (actual values), the ratio of time spent in the feeder in the light spot by normal conditions was calculated. Analysis of variance showed a main effect difference in genotype ($F(1,59)=4.138$) for time spent in feeder to be significant at $p<0.05$ despite of large error bars. This implicates that SCA17 spent less time in feeder compared to the wildtypes taking in consideration their behavioural response 24 hours before the light spot. We were unable to observed significant main effect differences in genotype & age at the progression of 9 months of the disease in other behavioural parameters; which include distance moved, total time and frequency in the shelter, time moving and not moving, velocity, frequency in feeder and time on shelter. There was an interesting observation when analyzing the variance of frequency on shelter where a main effect difference was seen in genotype by age interaction ($F(2,60)=5.324$, $p<0.01$). This indicates that the difference between genotype, to frequently jump on the shelter, depends on age. It is worth mentioning that this parameter, frequency to jump on shelter, is not anxiety-related but an interesting read-out for the general behaviour of the SCA17 rats and the progression of the disease. These results suggest an increased anxiety phenotype for the SCA17 rats seen when analysing the ratio of time spent before and on the day of the light spot test in the feeder. Findings from non-motor deficits of this rat model using the home cage monitoring will be presented in a descriptive manner and discussed.

Ethical statement

All experiments reported here were performed with the permission of the Animal Ethics Committee ('Lely-DEC') and in full compliance with the legal requirements of Dutch legislation on laboratory animals.

This study was supported by Marie Curie Initial Network (ITN) of the European Commission's 7th Framework Programme (PhenoRat).

References

1. A. C. Bruni, J. Takahashi-Fujigasaki, F. Maltecca, J. F. Foncin, A. Servadio, G. Casari, P. D'Adamo, R. Maletta, S. a M. Curcio, G. De Michele, A. Filla, K. H. El Hachimi, and C. Duyckaerts, "Behavioral disorder, dementia, ataxia, and rigidity in a large family with TATA box-binding protein mutation.," *Arch. Neurol.*, vol. 61, no. 8, pp. 1314–20, Aug. 2004.
2. Y. Wang, B. Bontempi, S. M. Hong, K. Mehta, P. R. Weinstein, G. M. Abrams, and J. Liu, "A comprehensive analysis of gait impairment after experimental stroke and the therapeutic effect of environmental enrichment in rats.," *J. Cereb. Blood Flow Metab.*, vol. 28, no. 12, pp. 1936–1950, 2008.
3. F. P. T. Hamers, A. J. Lankhorst, T. J. A. N. V. A. N. Laar, W. B. Veldhuis, and W. H. Gispen, "Automated Quantitative Gait Analysis During Overground Locomotion in the Rat : Its Application to Spinal Cord Contusion and Transection Injuries," vol. 18, no. 2, 2001.
4. C. Vandeputte, J.-M. Taymans, C. Casteels, F. Coun, Y. Ni, K. Van Laere, and V. Baekelandt, "Automated quantitative gait analysis in animal models of movement disorders.," *BMC Neurosci.*, vol. 11, p. 92, 2010.
5. E. Aarts, G. Maroteaux, M. Loos, B. Koopmans, J. Kovačević, A. B. Smit, M. Verhage, and S. Van Der Sluis, "The light spot test: measuring anxiety in mice in an automated home-cage environment," *Behav. Brain Res.*, vol. 294, pp. 123–130, 2015.
6. A. Rolfs, A. H. Koeppen, I. Bauer, P. Bauer, S. Buhlmann, H. Topka, L. Schöls, and O. Riess, "Clinical features and neuropathology of autosomal dominant spinocerebellar ataxia (SCA17).," *Ann. Neurol.*, vol. 54, no. 3, pp. 367–75, Sep. 2003.
7. G. Stevanin, A. Brice, M. Curie, S. Umr, and I. Fe, "Spinocerebellar ataxia 17 (SCA17) and Huntington's disease-like 4 (HDL4)," vol. 17, no. March, pp. 170–178, 2008.
8. K. Nakamura, S. Y. Jeong, T. Uchihara, M. Anno, K. Nagashima, T. Nagashima, S. Ikeda, S. Tsuji, and I. Kanazawa, "SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein.," *Hum. Mol. Genet.*, vol. 10, no. 14, pp. 1441–8, Jul. 2001.
9. P. Bauer, "Trinucleotide repeat expansion in SCA17/TBP in white patients with Huntington's disease-like phenotype," *J. Med. Genet.*, vol. 41, no. 3, pp. 230–232, Mar. 2004.
10. J. Hübner, a Sprenger, C. Klein, J. Hagenah, H. Rambold, C. Zühlke, D. Kömpf, a Rolfs, H. Kimmig, and C. Helmchen, "Eye movement abnormalities in spinocerebellar ataxia type 17 (SCA17).," *Neurology*, vol. 69, no. 11, pp. 1160–8, Sep. 2007.

Behavioural Assessment of a Transgenic Rat Model for Huntington's Disease – Limitations of Existing Tools for Rat Phenotyping

Giuseppe Manfré[1,2,3], Johanneke E. van der Harst[2], Judith R. Homberg[1] and Huu Phuc Nguyen[3]

[1]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands;
[2]Noldus Information Technology BV, Wageningen, The Netherlands
[3]Institute of Medical Genetics and Applied Genomics, University of Tuebingen, Tuebingen, Germany

All experiments reporter here were positively evaluated by the Animal Ethics Committee ('RU-DEC', Nijmegen) and performed under a project license from the Central Committee on Animals Experiments (CCD, The Hague), in full compliance with the legal requirements of Dutch legislation on the use and protection of laboratory animals.

The conventional behavioral phenotyping approach has been broadly used to investigate several aspects of animal behaviour. Typically, each animal is taken from the home cage and placed in the apparatus, where the test is carried out for a short period of time. The task is usually performed during either the light or the dark period and after the test the animal is returned to the home cage. Moreover, each test focuses solely on one aspect of behavior. Thus screening, for instance, a rodent model of a neurodegenerative disorder requires a series of different tests. Therefore, in order to investigate into detail the phenotypes preceding the onset of a certain disease classical behavioural paradigms need to be combined with novel automated and sophisticated methods which allow high-throughput testing.

Our research group focuses on Huntington's Disease (HD), an autosomal dominantly inherited progressive disorder which is characterized by psychiatric changes, dementia and motor dysfunction [1]. No effective treatment to influence the onset or the progression of this fatal disease is currently available. The behavioural characterization of a sensitive rodent model for HD is critical, especially in light of the development of therapeutic strategies for HD. In Tuebingen (Germany), our research group has recently generated an HD transgenic rat model using a human bacterial artificial chromosome (BAC), which contains the full-length HTT genomic sequence with 97 CAG/CAA repeats and all regulatory elements. BACHD transgenic rats display a robust, early onset and progressive HD-like phenotype including motor deficits and anxiety-related symptoms [2] rendering this model valuable for further phenotyping studies.

Although the onset of HD is clinically diagnosed on the basis of motor performance, symptoms of psychiatric disturbances such as anxiety, irritability, impulsivity, aggression, apathy and depressed mood are prevalent among pre-diagnostic HD gene carriers and patients with HD [3]. As many as 40–50% of patients with HD are found to experience depression, and depressed mood may precede disease onset by 4–10 years [4,5]. It is therefore important to recognize psychiatric symptoms in HD so that symptomatic treatment can be offered. This may be difficult later in the disease because diagnoses may be obscured by other features of the disease; depression, for instance, may be difficult to detect in a patient who has altered facial expressions and tone of voice. Conversely, metabolic symptoms such as weight loss and sleep disturbance may be wrongly attributed to depression. For these reasons it is of crucial importance to determine exactly when anxiety and depression appear in the course of HD and to determine time windows for future interventions.

The animals were therefore tested at different age points, to monitor the development of symptoms. The light-dark box and the elevated plus maze were used since these are validated and broadly used paradigms to assess depression and anxiety in rodent models [6-7]. But, as mentioned previously, carrying out classical behavioural paradigms is not enough to describe behaviour in detail. The need for improving throughput, validity, and reliability in the behavioral characterization of rodents may benefit from integrating automated intra-home-cage-screening systems allowing the simultaneous detection of multiple behavioral and physiological parameters in parallel. In order to do that, classical paradigms were combined with animal screening within phenotyping home-

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

cages (Phenotypers 4500), where several parameters were continuously measured. Anxiety was evaluated within the same environment challenging the animals with an automated task, the Light-Spot Test [8].

Male Sprague-Dawley rats (12 WT and 12 BACHD), were socially housed in group of two per cage in Rat Eurostandard Type-IV cages on a 12 hour light/dark cycle (lights on at 8 p.m.) with free access to food and water. On the first day of testing each animal was placed in a light-dark box for 10 minutes. The light/dark box test is based on the innate aversion of rodents to brightly illuminated areas and on the spontaneous exploratory behaviour of rodents in response to mild stressors, that is, novel environment and light. On Day 2 animals were housed for 7 consecutive days (from Day 2 to Day 8) in the Phenotypers (PT4500), an homecage environment that allows automated monitoring of animal behaviour [9]. On Day 7, the Light Spot test [8] took place inside the Phenotypers: a mild aversive stimulus was introduced fifteen minutes after the start of the dark phase (8:15 a.m.) for a period of 3 hours. This aversive stimulus consisted of a white light-spot shining from the upper left corner of the home-cage, with the light beam aiming at the feeding station. This test takes advantage of the natural tendency of rats to become highly active at the beginning of the dark phase and the association of this activity with feeding behavior. On Day 8 the animals were still housed in the Phenotypers and on Day 9 the animals were tested by using the Elevated Plus Maze (EPM). At the beginning of the test each animal was placed at the junction of the four arms of the maze, facing an open arm, and entries/duration in each arm were assessed. The detailed investigation of the behaviour of this transgenic rat model of HD is necessary for the early detection of the first core symptoms of the disease and the monitoring of its development so that possible therapies can be more effective in preventing further development of neurodegeneration.

References

1. Vonsattel JP, DiFiglia M (1998). Huntington disease. *Journal of Neuropathology & Experimental Neurology* 57:369 –384.
2. Yu-Taeger L, Petrasch-Parwez E, Osmand AP, Redensek A, Metzger S, Clemens LE, Park L, Howland D, Calaminus C, Gu X, Pichler B, Yang XW, Riess O, Nguyen HP (2012). A novel BACHD transgenic rat exhibits characteristic neuropathological features of Huntington disease. *The Journal of Neuroscience* 32(44):15426-38.
3. Rosenblatt A (2007). Neuropsychiatry of Huntington's disease. *Dialogues Clin Neurosci*. 9:191–197.
4. Pflanz S, Besson JA, Ebmeier KP, Simpson S (1991). The clinical manifestation of mental disorder in Huntington's disease: a retrospective case record study of disease progression. *Acta Psychiatr. Scand.* 83: 53–60.
5. Kirkwood SC, Su JL, Conneally M, Foround T (2001). Progression of symptoms in the early and middle stages of Huntington Disease. *Arch. Neurol.* 58: 273–278.
6. Bourin M, Hascoët M. (2003) The mouse light/dark box test. *Eur. J. Pharmacol.* 463(1-3): 55–65.
7. Walf AA, Frye CA (2007) The use of the elevated plus maze as an assay of anxiety-related behavior in rodents. *Nat. Protoc*. 2: 322–328.
8. Aarts E, Maroteaux G, Loos M, Koopmans B, Kovačević J, Smit AB, Verhage M, van der Sluis S, The Neuro-BSIK Mouse Phenomics Consortium (2015). The light spot test: measuring anxiety in mice in an automated home-cage environment. *Behav. Brain Res.* 294: 123–130.
9. de Heer RC, Schenke M, Kuurman WW, Spruijt (2008). Learning (in) the PhenoTyper: an integrative approach to conducting cognitive behavioural challenges in a home cage environment. *Proceedings of Measuring Behavior 2008 (Maastricht, The Netherlands, August 26-29, 2008)*. 57.

How to measure repetitive patterns of behavior: A translational behavioral genetic approach to Autism Spectrum Disorders

R.T. Molenhuis[1], H. Bruining[1,2], P.E. van Soldt[1], Myrna Brandt[1], F.A. Iraqi[3], R. Mott[4] and M.J.H. Kas[1]

[1]Department of Translational Neuroscience, Brain Center Rudolf Magnus, UMC Utrecht, The Netherlands
r.t.molenhuis-2@umcutrecht.nl
[2]Department of Psychiatry, Brain Center Rudolf Magnus, UMC Utrecht, The Netherlands
[3]Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Israel
[4]UCL Genetics Institute, University College London, UK

Background

Repetitive behaviors are one of the predominant features of Autism Spectrum Disorder (ASD). The term repetitive behavior is an umbrella term used to refer to the broad class of behaviors linked by repetition, rigidity, invariance and inappropriateness [1]. As there are currently no etiology-based treatments for repetitive behaviors in ASD, there is a need for behavioral assays to obtain translational preclinical models. Ilan Golani and colleagues have articulated that behavioral repertoires in animals can range from rich and free behavior on the one hand, to dull and predictable behavior on the other [2]. In this study, we aim to develop algorithms to detect and quantify repetitiveness in behavioral repertoires under novelty and baseline conditions, in different animal models relevant to ASD.

Methods

Test-environments and mouse models

To assess repetitiveness in behavioral repertoires under novelty conditions, mice are subjected to a single-trial exposure to four novel objects [3]. Exploration of these objects is scored manually using The Observer software (Noldus, Wageningen), and loco-motor activity is tracked using Ethovision software (Noldus, Wageningen). For the analysis of spontaneous behavioral repertoires under baseline conditions, mice are subjected to a multi-day home-cage experiment in the Noldus Phenotyper and tracked using Ethovision software (Noldus, Wageningen) [4].

C57BL/6J (C57) mice are contrasted with the BTBR T+tf/J (BTBR) mouse inbred strain, which is well-known for behavioral traits with face validity to all diagnostic symptoms of ASD [5]. Moreover, we have recently detected that BTBR mice contrast with C57 mice in novelty-induced exploratory behaviors [4]. Here we use BTBR mice as a tool to study repetitiveness in behavioral repertoires, quantified by various novel approaches.

Route-tracing stereotypies

Route-tracing stereotypies (i.e. repetitive patterns in loco-motor paths) are quantified by t-pattern sequential analysis as developed by Bonasera and colleagues [6]. T-pattern analysis reveals patterns by testing whether particular sequences of defined states occur within a specific time interval at a probability greater than chance. Analyses are performed using Theme software (Noldus, Wageningen). Correction for activity is applied by limiting the distance of the loco-motor path that is subjected to the t-pattern sequential analysis. We also compare the time-course of loco-motor activity.

Temporal patterning of object exploration

Repetitive patterns of novel object exploration are compared using t-pattern sequential analysis, which is performed using Theme software (Noldus, Wageningen). In this case, correction for activity is applied by

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

limiting the number of object-visits subjected to the t-pattern sequential analysis. Here we also compare the time-course of object-exploration activity.

Spatial distribution during novelty-exploration

To analyze richness in an animals' spatial distribution, the distribution of x- and y-coordinates is assessed during novel object exploration. Next, we characterize bouts of loco-motor activity during novel object exploration by duration and frequency. Finally, we assess the spatial distribution of bouts of activity during novel-object exploration.

Application of novel phenotypes to genome-wide analyses

To further understand the genetic architecture underlying our newly defined repetitive behavioral phenotypes, our novel quantitative measures were subjected to a genome-wide association study in >300 mic from the Collaborative Cross mouse genetic reference panel [7]. Multiple quantitative trait loci (QTL) were identified. We finally aim to assess association of known ASD risk-genes with these novel measures, to establish translational relationships between genotypes and repetitive behavior phenotypes [8].

Ethical statement

All experiments considering C57 and BTBR mice were approved by the ethical committee for animal experimentation of the University Medical Center Utrecht and performed according to the institutional guidelines that are in full compliance with the European Council Directive (86/609/EEC). All animal work on the Collaborative Cross was performed at the small animal facility at The Sackler Faculty of Medicine, Tel Aviv University, Israel. The Institutional Animal Care and Use Committee of Tel Aviv University approved all experimental protocols.

References

1. Turner M: Annotation: Repetitive behaviour in autism: a review of psychological research. *J Child Psychol Psychiatry* 1999, 40:839–49.
2. Golani I, Kafkafi N, Drai D: Phenotyping stereotypic behaviour: collective variables, range of variation and predictability. *Appl Anim Behav Sci* 1999, 65:191–220.
3. Pearson BL, Pobbe RLH, Defensor EB, Oasay L, Bolivar VJ, Blanchard DC, Blanchard RJ: Motor and cognitive stereotypies in the BTBR T+tf/J mouse model of autism. *Genes Brain Behav* 2011, 10:228–35.
4. Molenhuis RT, de Visser L, Bruining H, Kas MJ: Enhancing the value of psychiatric mouse models; differential expression of developmental behavioral and cognitive profiles in four inbred strains of mice. *Eur Neuropsychopharmacol* 2014:1–10.
5. Silverman JL, Yang M, Lord C, Crawley JN: Behavioural phenotyping assays for mouse models of autism. *Nat Rev Neurosci* 2010, 11:490–502.
6. Bonasera SJ, Schenk a K, Luxenberg EJ, Tecott LH: A novel method for automatic quantification of psychostimulant-evoked route-tracing stereotypy: application to Mus musculus. *Psychopharmacology (Berl)* 2008, 196:591–602.
7. Chesler EJ: Out of the bottleneck: the Diversity Outcross and Collaborative Cross mouse populations in behavioral genetics research. *Mamm Genome* 2013.
8. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al.: Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 2015, 87:1215–1233.

Impulsive Aggression: the Interrelationship Between Traits, SSRI Efficacy and 5-HT_{1A} Receptor Binding

D.G.A. Peeters[1,2], J. Rietdijk[1], N. Kogias[1], R.J. Verkes[2*] and J.R. Homberg[1*]

[1]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Nijmegen
[2]Department of Psychiatry, Radboud University Medical Centre, Nijmegen
***Both authors contributed equally to this work**

Background

The impact of the serotonergic system on aggression has been extensively studied, but our understanding remains inconclusive. Initial findings showed an association between reduced serotonin levels and aggression across species [1,2]. Therefore, selective serotonin reuptake inhibitors (SSRIs) seem to be good therapeutic candidates to increase extracellular serotonin levels and thereby decrease aggression. However, only a subset of aggressive patients seem to benefit from SSRI treatment. This subset might be defined by both behavioral and biological individual traits.

Aggressive behavior is not a unitary construct, but defined by the subjects environment and motivation. Premeditated aggression is goal-directed, whereas impulsive aggression reflects an aggressive response to environmental triggers with high levels of arousal. Most psychiatric patients suffer from impulsive aggression, which can be subtyped by several behavioral traits. Avoidance to social threats, anxiety and impulsivity are all associated with aggressive behavior and the serotonergic system. Furthermore, it has been shown that aggressive individuals scoring high on neuroticism showed an increased reduction in their aggressive behavior following SSRI treatment [3]. It is possible that other behavioral traits may also be predictive for the effectiveness of SSRIs as anti-aggressive drugs.

The mechanism by which SSRIs decrease aggression might also be relevant in predicting their effectiveness. It is well known that there is a delayed therapeutic effect of SSRIs, which might be caused by desensitization of 5-HT_{1A} receptors. These receptors seem to be an important therapeutic target, and numerous examples are known of strong anti-aggressive 5-HT_{1A} receptor agonists. Paradoxically, these agonists inhibit serotonin release, by activating the negative feedback loop in presynaptic neurons. However, these receptors are also found post-synaptically, where the agonist mimics higher serotonergic activity, corresponding to increased serotonin levels [4]. Functionally 5-HT_{1A} receptor binding potential has been shown to be decreased in the frontal cortex and amygdale of high aggressive individuals [5]. Altogether, individual differences in 5-HT_{1A} receptors and their functionality might also be important in predicting how aggressive individuals respond to SSRIs.

In the present study we try to elucidate whether behavioral traits and 5-HT_{1A} receptor binding are correlated with aggression and if they are able to predict the anti-aggressive efficiency of SSRIs.

Methods

Subjects and housing
All experiments were conducted in compliance with national regulatory principles and approved by the Committee for Animal Experiments of the Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. All efforts were made to reduce animal suffering and the number of experimental animals.

Male Long Evans rats (PND 100-110) were the experimental animals used as residents, housed in Macrolon type III cages, except during the resident intruder paradigm. Long Evans females (PND100-110) were sterilized by ligation of the oviducts, to keep them hormonally intact and housed socially (two per cage) in Macrolon type III cages when not housed together with a resident. Wistar males (PND 80-90) were used as intruders and housed

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

socially (four per cage) in Macrolon type IV cages.
All animals were housed on a 12 hour reversed light-dark cycle (lights on at 19:30, lights off at 7:30) in a temperature (21±1°C) and humidity-controlled room (60%) room. Animals had *ad libitum* access to water and food.

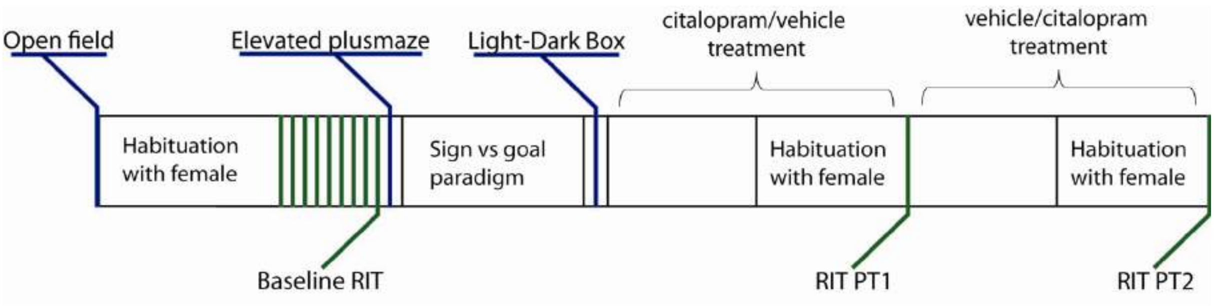


Figure 1: Experimental timeline showing behavioral experiments (RIT: resident intruder test, PT: post treatment).

Experimental design
Residents were repeatedly assessed for their anxiety levels with three distinct but comparable set ups: open field (OF), elevated plusmaze (EPM) and light-dark box (LDB). The open field test was performed before the first resident intruder interaction, the elevated plusmaze directly after the baseline aggression measurement and finally the light dark box the day before treatment started (see figure 1). In the elevated plusmaze and light-dark box tests animals were able to freely explore the apparatus for 5 min, whereas in the open field this was extended to 15 min. Movement of the animals was tracked by Ethovision 9 (Noldus). 'Time spent in' and 'latency to first entry of' the centre zone (OF), open arms (EPM) and light compartment (LDB) were compared between experiments.
In between the elevated plusmaze and light-dark box tests a sign versus goal paradigm was used to evaluate differences in the degree to which residents attribute incentive salience to reward predicting cues.
After the light-dark box, animals were chronically treated with either citalopram hydrobromide (Sigma-Aldrich) (20 mg/kg; dissolved in saline) or saline, using osmotic minipumps (2ML2 Alzet, USA, 5 ul/hr, 14 days) which were implanted subcutaneously under isoflurane anesthesia. A crossover design was used in which every animal received both treatments, and was tested after every 14 days (see figure 1).
Aggressive behavior was evaluated with a resident intruder paradigm, as described below.

Resident intruder paradigm
The resident intruder paradigm was based on the protocol published by Koolhaas and colleagues [6]. One week before the start of the interactions each resident was housed in a PhenoTyper 4500 cage (45x45x55 cm, Noldus), since the interaction requires sufficient space to be able to observe the full range of behaviors. To facilitate territorial behavior and prevent social isolation, an hormonally intact sterilized female was housed together with the resident starting one week before the first interaction and ending on the day of the final resident intruder interaction. Bedding material was not cleaned during the entire testing period, since territoriality is strongly based on olfactory cues.
Building up to the baseline measurement, testing occurred once every other day. On every test day the female was removed one hour before the test took place. At the start of the test an unfamiliar, slightly smaller Wistar male was introduced in the cage of the resident as the intruder. The first eight test days the males were allowed to interact for ten minutes and separated after the residents' first clinch attack. Only on the ninth test day the males were allowed to interact a full ten minutes after the first clinch attack. The level of aggression during this 10 minute interaction was considered the baseline aggression. After completion of the test, the females were reunited with the residents and the intruders were returned to their homecage.
Post treatment aggression levels were determined after one week habituation with a female as described in a single full ten minute interaction after the first clinch attack.
Behavior was videotaped and afterwards scored using Observer 11 (Noldus). On every test day attack latencies were determined. During the full ten minute interactions duration and frequency of the following offensive

behaviors shown by the resident were determined: move towards, chase, social exploration, lateral threat, clinch attack, keep down, upright posture, dragging and inactivity.

Autoradiography
After completing the behavioral tests resident brains will be prepared for 5-HT_{1A} autoradiography to determine 5-HT_{1A} binding.

Results and expectations

Preliminary results suggest a negative correlation between anxiety and baseline aggression, with robust and comparable results from open field and elevated plus maze data. Citalopram treatment will most likely have the strongest anti-aggressive effect in the highest aggressive animals. Behavioral data and 5-HT_{1A} receptor binding will be correlated to aggression levels post treatment to determine their predictive value in SSRI responsiveness.

References

1. Rosell, D.R., and Siever, L.J. (2015) The neurobiology of aggression and violence. CNS Spectrums, Available on CJO 2015 doi:10.1017/S109285291500019X
2. Duke, A.A., Bégue, L., Bell, R., EisenIohr-Moul, T. (2012) Revisiting the Serotonin–Aggression Relation in Humans: A Meta-Analysis. Psycholoigical Bulletin Volume 139(5), 1148-1172
3. Phan KL, Lee R, Coccaro EF. (2011) Personality predictors of antiaggressive response to fluoxetine: inverse association with neuroticism and harm avoidance. Int Clin Psychopharmacol. 26(5):278-8
4. De Boer, S.F., Newman-Tancredi, A. Anti-aggressive effects of the selective high-efficacy ‘biased’ 5-HT_{1A} receptor agonists F15599 and F13714 in male WTG rats. Psychopharmacology 1-11
5. Popova, N.K. (2006) From genes to aggressive behavior: the role of serotonergic system. BioEssays, 28, 495-503
6. Koolhaas, J.M., Coppens C.M., De Boer, S.F., Buwalda B., Meerlo P., Timmermans P.J.A. (2013) The resident-intruder paradigm: a standardized test for aggression, violence and social stress. J. Vis Exp. (77), e4367

The Role of the Nigrostriatal Dopamine System in Cognitive Function

M.J. Lelos

School of Biosciences, Cardiff University, Cardiff, Wales, UK. lelosmj@cardiff.ac.uk

One of the key neuropathological hallmarks of Parkinson's disease is the degeneration of the nigrostriatal pathway and the consequent denervation of dopaminergic fibres to the striatum. While the cardinal symptoms of Parkinson's disease are overt motor impairments, a wide range of non-motor symptoms also manifest. Recent studies have revealed that these non-motor symptoms can impact upon the quality of life for people with Parkinson's disease even more than the motor complications. Despite this, little is known about the pathology underlying these non-motor symptoms, although non-dopaminergic origins are often assumed.

To address this question, our aim was to first investigate whether depletion of dopamine from specific striatal subregions could underlie the manifestation of cognitive or neuropsychiatric non-motor symptoms. In Experiment 1, rats received bilateral lesions of the dorsomedial striatum that damaged either the medium spiny projection neurons (with quinolinic acid) or the terminal dopaminergic fibres (with 6-hydroxydopamine). Rats were tested in automated 2-lever operant boxes on the delayed alternation task, a test of working memory. The results revealed significantly impaired performance in rats with both types of striatal damage, revealing for the first time a role of medial striatal dopamine in performance in this task. To determine whether this effect was specific to the medial region, Experiment 2 compared rats with the same dorsomedial striatal dopamine lesions with rats receiving terminal lesions of the dorsolateral striatum. The results revealed a specific impairment in rats with dorsomedial striatal dopamine loss and no effect of dopamine depletion in the lateral subregion. Analysis of the results suggested that the deficit was not related to the working memory component of the task, but rather to the ability to utilize the alternation rule. A probe test was conducted in the T-maze to verify that spontaneous alternation performance remained intact in rats with dorsomedial striatal dopamine depletion. Subsequent exposure to chronic levodopa in these rats was found to have no impact upon performance in the task, although the development of dyskinesias was evident in the laterally lesioned rats, but not those with medial lesions. Thus, our data reveal a dissociation in the roles of medial and lateral striatal dopamine transmission in cognitive processing and dyskinesia development, respectively.

Other non-motor symptoms of Parkinson's disease include changes in visuospatial processing, attentional impairments and neuropsychiatric symptoms. To address whether deterioration of these non-motor functions may be attributed directly to degeneration of the nigrostriatal pathway, the next experiment was conducted using the unilateral 6-hydroxydopamine lesion model of Parkinson's disease. Rats were trained for 6 weeks on the lateralized choice reaction time task in 9-hole operant boxes. A subcohort of rats then received unilateral lesions of the medial forebrain bundle, resulting in loss of dopamine from the right hemisphere. Drug-induced rotation tests using meth-amphetamine challenge (2.5mg/kg) revealed losses of ≥95% dopamine in the striatum in this cohort. Analysis of behavioural performance on the lateralized choice reaction time task revealed significant impairments in motor function (movement time), visuospatial function (accuracy), attentional processing (reaction time) and motivation (useable trials) in lesioned rats. A subset of dopamine-depleted rats was then grafted with dopamine-rich transplants of human ventral mesencephalic tissue into the right striatum. At 18 weeks post-graft, rats grafted with human dopaminergic tissue displayed improvements in drug-induced rotational bias, motor function, visuospatial performance and motivation, but not in attentional function. Thus these data demonstrate the role of nigrostriatal dopamine in motor, as well as non-motor, function in a rodent model of Parkinson's disease. Moreover, these data also provide the first evidence of improvements in non-motor dysfunctions (cognitive and neuropsychiatric impairments) using the transplantation of human primary dopamine cells as a therapeutic intervention.

Overall, these experiments have revealed 1) a novel role for dorsomedial striatal dopamine transmission in rule learning, 2) the necessity for loss of dorsolateral striatal dopamine for the emergence of levodopa-induced dyskinesias and 3) that deterioration of the nigrostriatal pathway induces a range of non-motor impairments and

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

improvements in visuospatial processing and motivation can be observed after the striatal transplantation of human primary dopamine cells.

All experiments were conducted in compliance with the UK Animals (Scientific Procedures) Act 1986 under Home Office Licence No. 30⁄2498 and with the approval of the local Cardiff University Ethics Review Committee.

Behavioral Phenotyping of Rodent Models for Neuropsychiatric Disorders: Effects of Post-weaning Social Isolation

D. Seffer, R.K.W. Schwarting and M. Wöhr

Behavioral Neuroscience, Experimental and Physiological Psychology, Philipps-University of Marburg, Marburg, Germany. seffer@staff.uni-marburg.de

Abstract

Rats are highly social animals and rough-and-tumble play during adolescence has an important role for social development. Post-weaning social isolation, i.e. separation from conspecifics during this phase, is widely used to study the adverse effects of juvenile social deprivation and to induce behavioral phenotypes and changes in neural development relevant to neuropsychiatric disorders like autism and schizophrenia. Ultrasonic vocalizations (USV) are an important component of the rat's social behavioral repertoire and serve as situation-dependent affective signals with important communicative functions. We found that post-weaning social isolation impairs approach behavior in response to pro-social USV in our 50-kHz USV radial maze playback paradigm. Besides ultrasonic communication deficits, post-weaning social isolation also led to cognitive impairments as indicated by poor novel object recognition. At the neurobiological level, post-weaning isolation led to alterations in the microRNA-dependent Ube3a1 function; a key regulator of neuronal development and synaptic plasticity. Taken together, these results highlight the relevance of post-weaning social isolation as a rodent model of neuropsychiatric disorders with impaired social communication like autism and schizophrenia allowing the investigation of underlying genetic and neurobiological alterations.

Introduction

Rats are highly social animals and live in groups with a distinct hierarchy, offering optimal experimental conditions to investigate social behavior and communication. Rough-and-tumble play during adolescence has an important role for social development. Post-weaning social isolation, i.e. separation from conspecifics during this phase, leads to a behavioral phenotype characterized by prominent social impairments in adulthood [1; 2]. Most notably, long-term changes in behavior with relevance to various neuropsychiatric disorders can be observed, including impaired sensorimotor gating, cognitive inflexibility, increased aggressive behavior, and social withdrawal [3-5]. These behavioral alterations are further accompanied by various neuromorphological changes and neurochemical imbalances, including neurotransmitter systems implicated in neuropsychiatric disorders, such as dopamine, glutamate and serotonin.

Ultrasonic vocalizations (USV) are an important component of the rat's social behavioral repertoire and serve as situation-dependent affective signals with important communicative functions [6-8]. High-frequency 50-kHz USV are produced in appetitive situations such as rough-and-tumble play and induce social approach behavior, indicating that they serve as social contact calls [9]. We recently demonstrated that playback of 50-kHz USV consistently led to social approach behavior in the recipient, indicating that pro-social ultrasonic communication can be studied in a reliable and highly standardized manner by means of the 50-kHz USV radial maze playback paradigm [10].

Methods

Animals: Male Wistar rats (HsdCpb:WU, Harlan, Venray, The Netherlands) served as subjects and were housed in an animal room with a 12:12 h light/dark cycle (lights on 8–20 h) where the environmental temperature was maintained between 20 and 23°C (humidity: 30–50%). Lab chow (Altromin, Lage, Germany) and water (0.0004% HCl-solution) were available *ad libitum*. Rats were handled for three consecutive days prior to testing in a standardized way for 5min. All experimental procedures were performed according to legal requirements of

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Germany and approved by the ethical committee of the local government (Regierungspräsidium Gießen, Germany).

Post-weaning Social Isolation: Rats were weaned at about 3 weeks of age and subsequently housed in one of the following conditions: a) group housing (NO ISO), with rats being housed in groups of six; b) short-term social isolation (SHORT ISO), with rats housed in groups of six, but isolated 24 h prior testing; or c) long-term social isolation (LONG ISO), with rats being isolated for 28 days prior testing (see Figure 1). At about 7 weeks of age they were tested in the 50-kHz USV radial maze playback paradigm.

Post-weaning Social Isolation and Re-socialization: Since recent findings demonstrated that social deficits displayed by a well-established autism model can be improved by peer intervention [11], we tested for phenotypic rescue by adding one additional week of peer rearing after being housed under one of the three experimental housing conditions for 4 weeks, i.e., a) NO ISO, b) SHORT ISO, or c) LONG ISO. Consequently, rats were tested at about 8 weeks of age in the 50-kHz USV radial maze playback paradigm. Of note: For re-socialization the previous peers were re-united.

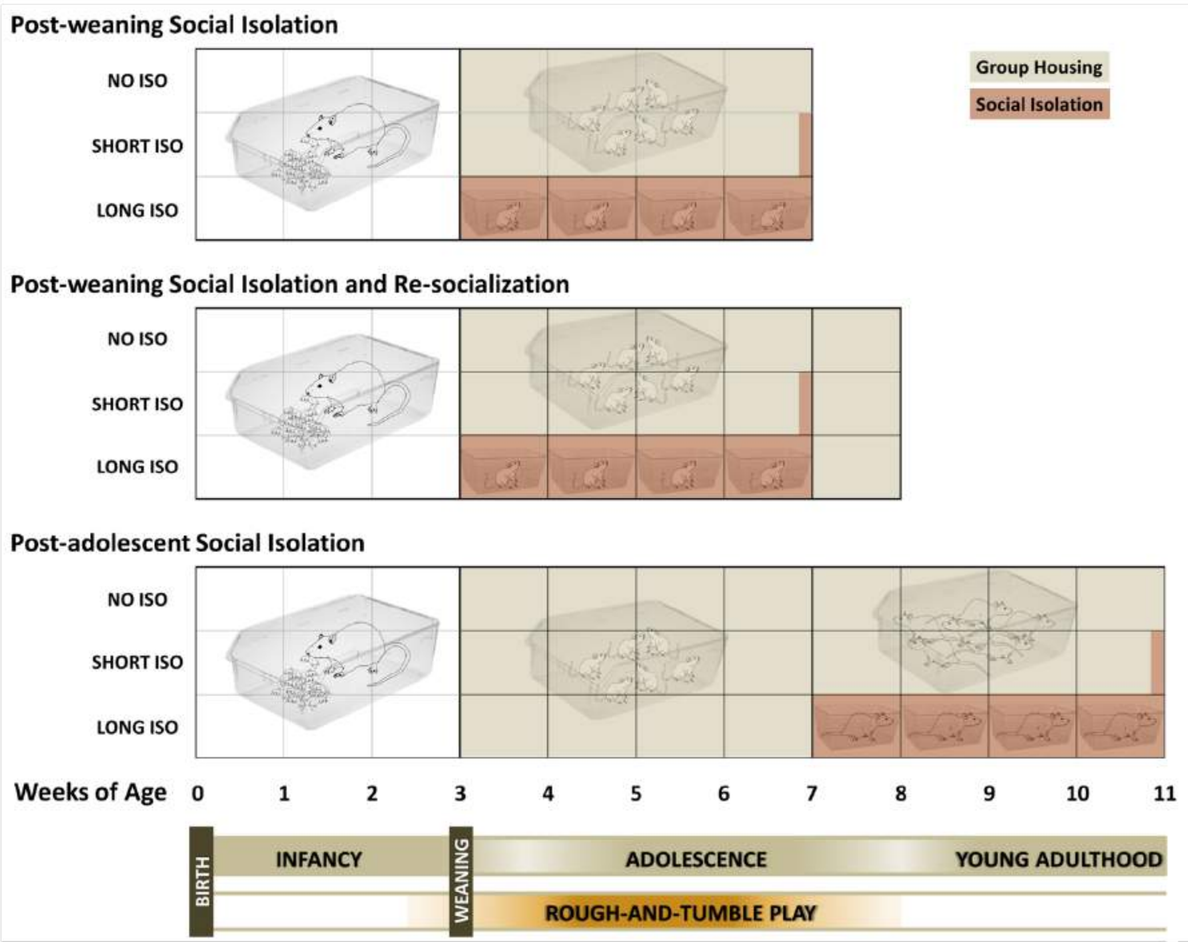


Figure 1. Top) Experimental design to assess the impact of social isolation on approach behavior induced by pro-social 50-kHz USV. Experimental designs: Post-weaning social isolation; Post-weaning social isolation and re-socialization; and Post-adolescent social isolation. Bottom) Schematic illustration of the rat developmental profile, highlighting the rough-and-tumble play period.

Post-adolescent Social Isolation: We tested if isolation-induced deficits depend on the time period of isolation during development. To this aim, post-adolescent young adult rats at about 7 weeks of age, and after going through the rough-and-tumble play period [12], were exposed to one of the three experimental housing conditions for 4 weeks, i.e., a) NO ISO, b) SHORT ISO, or c) LONG ISO. Consequently, rats were tested at about 11 weeks of age in the 50-kHz USV radial maze playback paradigm.

Housing: Group housing was conducted in polycarbonate Macrolon type IV cages (380 × 200 × 590mm, plus high stainless steel covers) and isolation rearing in Macrolon type III cages (265 × 150 × 425mm, plus high stainless steel covers), both filled with Tapvei peeled aspen bedding (indulab ag, Gams, Switzerland).

Experimental setup: Testing was performed in the 50-kHz USV radial maze playback paradigm (see Figure 2, for details see: [10; 13]). Prior to each test, behavioral equipment was cleaned using a 0.1 % acetic acid solution followed by drying.

Experimental procedure: A given animal was placed onto the central platform of the radial maze, facing the arm opposite to the loudspeaker. After an initial phase of 15 min where no acoustic stimuli were presented (habituation), the rat was exposed to one of the following three acoustic stimuli for 5min: a) 50-kHz USV; b) 22-kHz USV; or c) background noise with the latter two serving as control stimuli. After a post-stimulus interval of 60 min, the subjects were deeply anaesthetized and their brains were removed.

Behavioral analysis: Behavioral analysis was performed using an automated video tracking system (EthoVision, Noldus Information Technology, Wageningen, The Netherlands, see Figure 2). Previously, the following parameters were obtained: 1) total distance travelled (cm), 2) number of arm entries into the three

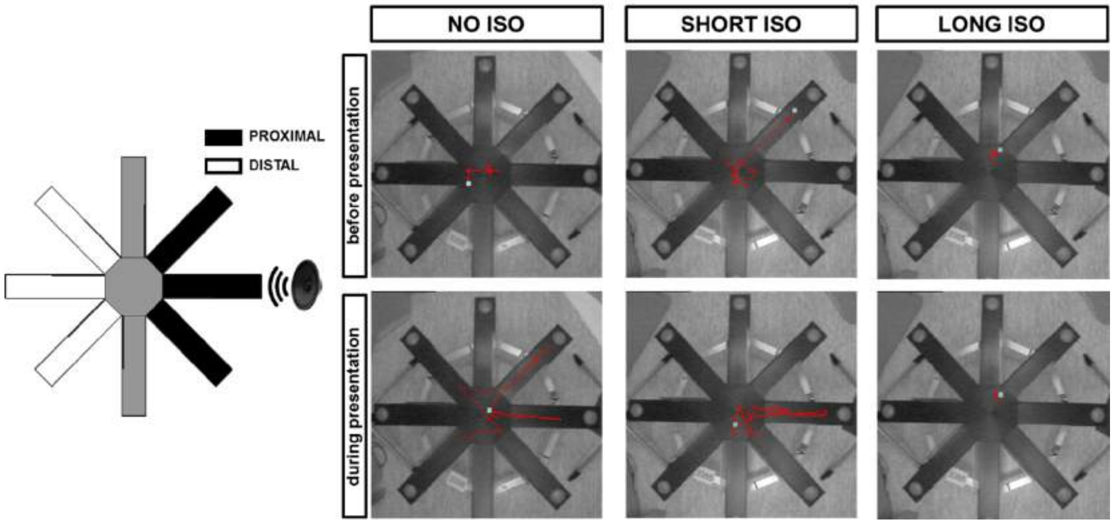


Figure 2. Left) Schematic illustration of the elevated radial eight arm maze. Arms close to the loudspeaker are denoted as proximal, central arms as neutral and opposite arms as distal. Right) Exemplary patterns of a rat's behavioral activity after either group-housing (NO ISO), post-weaning short-term social isolation (SHORT ISO) or post-weaning long-term social isolation (LONG ISO) before and during presentation of pro-social 50-kHz USV measured by using the automated video tracking software EthoVision (Noldus Information Technology, Wageningen, The Netherlands), with tracking profiles shown in red.

proximal or distal arms (n) and 3) the time spent thereon. Here, the track profile was acquired. For the automated analysis, input filters were activated to avoid an over-estimation of locomotor activity due to head-movements: a minimal distance moved of 8 cm was used for the total distance travelled, whereas a minimal distance moved of 3 cm was used for the arm entries.

Results

Post-weaning social isolation specifically affected the behavioral response to playback of pro-social 50-kHz. While group-housed rats showed the expected preference toward 50-kHz USV, the response was even stronger in short-term isolated rats, possibly due to a higher level of social motivation (see Figure 2). In contrast, post-weaning long-term isolation led to pronounced deficits, with rats rather displaying avoidance behavior. Importantly, impaired approach behavior after post-weaning social isolation was rescued by peer-mediated re-socialization, and isolation-induced deficits depend on the time period of isolation during development [13].

At the neurobiological level, post-weaning isolation, also resulting in poor novel object recognition as expected, led to an increase in an alternative E3 ubiquitin ligase Ube3a transcript, Ube3a1, in the hippocampus; a key regulator of activity-dependent synapse development and plasticity [14]. The increase in Ube3a1 RNA expression following post-weaning isolation was paralleled by elevated levels of microRNA 134, with Ube3a1 knockdown increasing dendritic complexity in the hippocampus in wild-type controls. Ube3a1 RNA knockdown, however, failed to induce dendritic complexity when the miRNA cluster 379-410, including miR-134, was missing, demonstrating that the Ube3a1 function is microRNA-dependent.

Discussion

These results indicate a critical involvement of play behavior for the social development during adolescence in rats. Post-weaning social isolation led to ultrasonic communication deficits, cognitive impairments and alterations in microRNA-dependent Ube3a1 function on neuronal plasticity. Deficits in the social behavior repertoire mimic negative symptoms typically observed in schizophrenia. The finding that environmental factors affecting social behavior and cognition alter Ube3a has important implications, particularly since loss of UBE3A is the leading cause for the neurodevelopmental disorder Angelman syndrome and UBE3A duplications are among the most frequent copy number variations associated with autism. Taken together, these results highlight the relevance of post-weaning social isolation as a rodent model of neuropsychiatric disorders with impaired social communication like autism and schizophrenia allowing the investigation of underlying genetic and neurobiological alterations. This approach would likely lead to the identification of promising targets for novel pharmacological treatment approaches, the efficacy of which could be tested in this rodent model.

References

1. Hol, T., van den Berg, C.L., van Ree, J.M., Spruijt, B.M. (1999). Isolation during the play period in infancy decreases adult social interactions in rats. *Behavioural Brain Research* **100**, 91-97.
2. van den Berg, C.L., Hol, T., van Ree, J.M., Spruijt, B.M., Everts, H., Koolhaas, J.M. (1999). Play is indispensable for an adequate development of coping with social challenges in the rat. *Developmental Psychobiology* **34**, 129-138.
3. Fone, K., Porkess, V. (2008). Behavioral and neurochemical effects of post-weaning social isolation in rodents – Relevance to developmental neuropsychiatric disorders. *Neuroscience & Biobehavioral Reviews* **32**, 1087-1102.
4. Hall, F.S. (1998). Social deprivation of neonatal, adolescent, and adult rats has distinct neurochemical and behavioral consequences. *Critical Reviews in Neurobiology* **12**, 129-162.
5. Lapiz, M.D., Fulford, A., Muchimapura, S., Mason, R., Parker, T., Marsden, C.A. (2003). Influence of postweaning social isolation in the rat on brain development, conditioned behavior, and neurotransmission. *Neuroscience & Behavioral Physiology* **33**, 13-29.
6. Brudzynski, S. M. (2013). Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. *Current Opinion in Neurobiology* **23**, 310-317.
7. Portfors, C.V. (2007). Types and functions of ultrasonic vocalizations in laboratory rats and mice. *Journal of the American Association for Laboratory Animal Science* **46**, 28-34.
8. Wöhr, M., Schwarting, R.K.W. (2013). Affective communication in rodents: ultrasonic vocalizations as a

tool for research on emotion and motivation. *Cell and Tissue Research* **354**, 81-97.

9. Wöhr, M., Schwarting, R.K.W. (2007). Ultrasonic communication in rats: can playback of 50-kHz calls induce approach behavior? *PLoS ONE* **2**, e1365.
10. Seffer, D., Schwarting, R.K.W., Wöhr, M. (2014). Pro-social ultrasonic communication in rats: insights from playback studies. *Journal of Neuroscience Methods* **234**, 73-81.
11. Yang, M., Perry, K., Weber, M.D., Katz, A.M., Crawley, J.N. (2011). Social peers rescue autism-relevant sociability deficits in adolescent mice. *Autism Research* **4**, 17-27.
12. Panksepp, J. (1981). The ontogeny of play in rats. *Developmental Psychobiology* **14**, 327-332.
13. Seffer, D., Rippberger, H., Schwarting, R.K.W., Wöhr, M. (2015) Pro-social 50-kHz ultrasonic communication in rats: post-weaning but not post-adolescent social isolation leads to social impairments - phenotypic rescue by re-socialization. *Frontiers in Behavioral Neuroscience* **9**:102.
14. Valluy, J., Bicker, S., Aksoy-Aksel, A., Lackinger, M., Sumer, S., Fiore, R., Wüst, T., Seffer, D., Metge, F., Dieterich, C., Wöhr, M., Schwarting, R.K.W., Schratt, G. (2015). A coding-independent function

Observing and Classifying Mutual Gaze Patterns

Guido Ferdinand Lichtert

Research Group Experimental ORL
Department Neurosciences
Katholieke Universiteit Leuven, Belgium

Introduction

One of the most fundamental and powerful modes of interpersonal encounter is the interaction between two pairs of eyes and the energy that is mediated by this interaction [1]. From birth on, the infant is most receptive and responsive to the caretaker's eye signals, an encounter which sets a lively and lifelong social exchange [2, 3]. Parents often experience their newborn baby as a real human being when their eyes encounter each other [4]. They make the effort to get eye contact by trying to find the optimal focal distance from the child's eyes to their own eyes. This phenomenon was described by Papousek and Papousek [5] as 'visual cuing'. Mutual gaze is an important aspect of face to face communication and is influenced by the characteristics of the interaction partners [6]. Mutual gaze however becomes an essential aspect of communication when at least one of the interaction partners is deaf and does not have full access to auditory information transferred by the other. In this study we investigated the feasibility and reliability of coding mutual gaze by hearing mother - deaf child dyads. We also describe how mutual gaze patterns can be clearly defined and classified by these measures as separate entities

Subjects

Eighteen typical developing children who were profoundly deaf before speech development were examined longitudinally at 18, 24 and 30 months of age in a play interaction with their hearing mother. All children are congenitally deaf with a Fletcher Index on the better ear of at least 90 dB SPL. Above 90 dB hearing loss is the highest category. All children were using conventional hearing aids. None of these children had a cochlear implant at the moment of research which means that they had to rely on mutual gaze for communication.

Procedure

Recordings for the observations were made in a standardised 'laboratory' living room of a home guidance service that supports parents in raising their deaf child. This allows a good balance between a standardized and familiar environment. Both mother and child are seated on either side of the table with the child in a custom highchair. This position was specially chosen as the best setup to record clear mutual gaze because both persons have to move their head to look at each other. For the play interaction, stacking blocks were chosen as a toy set because they are a well-known play routine [7] with a strong power to elicit mutual gaze, structured attention [8] and turn taking, which are necessary skills to develop smooth human conversation. The video recordings of the play interaction were made with two visible and unmanned cameras to guarantee secure dyadic interactions. One camera was focused on the mother and the other on the child, with both subjects positioned on the left and right sides of the camera frame. Both images were merged in the control room to one split-screen image by an effect generator. The merged recordings are coded manually frame by frame by means of an ergonomic and custom made computer program directly linked to the video recorder [9]. Three types of gazes were distinguished: gazing at the face of the other, gazing at the toy and gazing at something else. These types of gaze were coded separately for the mother and child. For accurate measurement of the gaze direction, clear operationalizations were made for the different conditions of the head and eye movements, position of the eye pupil and eyelid (palpebra) and visibility of the sclera (see Table 1). The begin and end time of the gaze direction were stored in a database which automatically calculated the duration of these gaze parameters separately for mother and child as well as mutual gaze and co-orientation of both subjects together. Mutual gaze is the moment when both interactants look at each other's face, while co-orientation means that parent and child look simultaneously at the

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

same object. These data were plotted in separate but linked graphs for mother and child. A ruler was programmed as a tool to easily define and visualize gaze patterns. This is a red vertical bar that can moved manually over the plotted graphs of the visual behavior of child and parent to locate precisely the beginning and the end of the mutual gaze of both subjects.

To make a reliable analysis and to put the behavior in its natural context, we first looked at the whole fragment at normal speed before starting analyzing frame by frame.

Table 1. Decision rules for 'looking at' and 'looking away' from the face of the other

CONDITION	LOOKING AT	LOOKING AWAY
1. Palpebrae are moving		
1.1 Sclera is visible	From the time the white of the eyeball just appears	From the time the white of the eyeball just disappears
1.2.Sclera is invisible	From the time the eyes are just opened again	From the time the eyes are just closed again
2. Palpebrae are not moving		
2.1 Sclera is visible		
Pupil comes fully to one corner of the eye	The white of the eyeball in one eye is fully visible alongside the pupil in the eye corner	The same as 'looking at' but in the opposite direction
Pupil does not come fully to one corner of the eye	From the time that the white is visible from the corner of the eye from which the pupil moves away	From the moment that the pupil moves back to the corner of the eye
2.2. Sclera is invisable	from the first global head or body movement to the face of the other	from the first global head or body movement away from the face of the other

Results

Based on inter- and intra-scorer reliability measures we concluded that gazing can be coded reliably. Cohen kappa's for inter-scorer reliability are 0.93 and 0.89 for child and mother, respectively, for all 3 ages and 0.96 and 0.82 respectively for the intra-scorer reliability. Gazing can be coded especially precisely in children. It appears to be more difficult, though still reliable, to code gazing in adults because the gazes are often shorter and the visual direction is less clearly focused than those of children. Adults seem to use a broader gaze angle to focus on more things at the same time. During the analysis, it was striking how the change of gaze direction, particularly in children, involved closing and opening of the eyes. This phenomenon was already noted in the study of de Leuw in 1988 [10]. The movements of the palpebrae make it easy to decide when the child was looking at or looking away from the face of the adult, based on the white of the sclera that appears or disappears. In the mothers however, the eyes stayed often open by quickly changing gaze directions. In such cases, decisions about the gaze direction were derived from the movement of the eyeball.

Accurate measuring of gaze directions of both agents also allows clearly defining, labeling and classifying mutual gaze patterns as separate entities. The two main categories are solitary and dyadic gazing. Solitary gazing is defined as a gaze contact of an adult or a child which is not replied by a mutual gaze of the other within an interaction loop. A dyadic gazing pattern is defined as a series of gazing patterns by the adult and the child, in which both partners sequentially have initiated and concluded their gazing pattern. For the labeling of these patterns a notation system was developed which allows constructing a taxonomy of gaze patterns according to number of participants, complexity of gaze movements, and onset of interaction and timing of gazing. Longitudinal research has also proved the construct validity of these patterns. Construct validity is the appropriateness of inferences made on the basis of observations or measurements (often test scores), specifically whether a test measures the intended construct. The measuring of these gaze patterns have proven to be very useful in the study of deaf toddlers in interaction with their hearing mothers. These different patterns were highly predictive for language outcomes at a later age. This method has been successfully used to show that there was a difference in mother-child visual communication interactions between children with advanced hearing technology (cochlear implants) and children without cochlear implants. These measures are also helpful to understand the development of joint attention from typical developing children or children with different disorders such as autism spectrum disorders, epileptic disorders or depressions, and its influence on other developmental tasks like cognition, theory of mind, emotional development, etc. Every study that focusses on mutual visual turn taking during conversations can use these coding schemes to give a more detailed and accurate analysis of this process.

Acknowledgments

This work is supported by the FP7 people program (Marie Curie Actions), REA grant agreement no FP7-607139 (iCARE).

References

1. Angus, N., Osborne, J., & Koziey, P. (1991). Window to the Soul: A phenomenological investigation of mutual gaze. *Journal of Phenomenological Psychology, 22*, 142-162.
2. Brazelton, T.B., & Cramer, B.G. (1990). *De eerste band. Ouders, kinderen en hun vroegste binding* (Vert., S. Wagenaar, The earliest relationship, 1990). Baarn: Bosch & Keuning.
3. Riess, A. (1988). The power of the eye in nature, nurture and culture. A developmental view of mutual gaze. *Psychoanalytic Study of the Child, 43,* 399-421.
4. Robson, K.S. (1967). The role of eye-to-eye contact in maternal-infant attachment. *Journal of Child Psychology and Psychiatry, 8,* 13-25.
5. Papousek, H., & Papousek, M. (1987). Intuitive parenting: A dialectic counterpart to the infant's integrative competence. In J.D. Osofsky (Ed.), *Handbook of infant development* (2nd edition, pp. 669-720). NewYork: John Wiley & Sons.
6. Broz, F., Lehmann, H.,Nehaniv, C.L., & Dautenhahn, K. (2012). *Mutual Gaze, Personality, and Familiarity: Dual Eye-Tracking During Conversation.* The 21st IEEE International Symposium on Robot and Human Interactive Communication, September 9-13, Paris, France,.858-864
7. Bruner, J.S. (1983) *Child's talk. Learning to use language*. NY, London, W.W. Norton & Company.
8. Wood, D., Wood, H., Griffiths, A., & Howarth, I. (with Tait, M. & Lewis, S.) (1986). *Teaching and talking with deaf children*. Chichester, New York: John Wiley & Sons.
9. Lichtert, G. (1999). *Onderzoek naar het kijkgedrag van horende moeders en hun dove peuter in relatie tot de communicatieve ontwikkeling van het kind.* [*Research into gaze behavior of hearing parents and their deaf toddler in relation to the communicative development of the child*]. Unpublished doctoral dissertation, Katholieke Universiteit Leuven, Belgium.
10. De Leuw, C. (1988). *Een observationeel onderzoek naar de effectiviteit van de communicatie tussen moeders en hun prelinguaal dove kinderen in een spelsituatie.* [*An observational study of the effectiveness of communication between mothers and their prelingual deaf children during play*]. Unpublished master thesis Katholieke Universiteit Nijmegen, Netherlands.

Auditory Environment across the Lifespan of Cochlear Implant Users: Insights from Data Logging

Tobias Busch[1,2], Filiep Vanpoucke[2], Astrid van Wieringen[1]

[1] Department of Neurosciences, Research Group Experimental ORL, KU Leuven, Leuven, Belgium.
tobias.busch@kuleuven.be
[2] Cochlear Technology Centre, Mechelen, Belgium.

Many profoundly deaf people are able to perceive speech by means of a cochlear implant (CI). A CI bypasses the damaged cochlea by directly stimulating the auditory nerve with electrical currents. For persons with a CI the auditory environment is crucial: Aspects like the quantity and clarity of speech and the amount of noise in the environment can affect treatment outcome and patient satisfaction. Yet, little is known about the sound environments that people with CI experience throughout the day.

CI data logging is a new method that allows unobtrusive, comprehensive monitoring of the natural acoustical environment of CI users. The Cochlear Nucleus 6 CI sound processor continuously analyzes the acoustical environment of the wearer and automatically classifies it as one of six scenes (quiet, speech, speech in noise, noise, wind, or music). The processor registers the time spent in each scene, the sound pressure level and other usage related information in a data log.

We did an exploratory retrospective cross-sectional analysis of a convenience sample of data logs. Our sample contained anonymized logs of more than 1200 CI users of all ages from 13 countries around the world, covering two million hours of implant use. We used hierarchical linear modeling to investigate differences between age groups and variability between users and countries. Specifically, we analyzed daily device use, the amount of time spent in different acoustical scenes, the sound pressure levels users were exposed to, and the quantity and quality of the speech input they received.

We found that the majority of CI recipients used their device for the greater part of the day. However, there was a substantial amount of variation and partial use. We also found age related differences in the auditory environment of CI users: There were changes in the amount of exposure to the different scenes, the sound intensity of the environment, and the quality of the speech input. Across all age groups, users spent a large part of their time in noisy environments. Accordingly, a lot of speech was presented in background noise. However, even within the same age group and geographical region we found large inter-individual variation for all measured variables.

This study was exploratory and our sample had considerable limitations. Nevertheless, our results shed light on the everyday acoustical experience of CI users. We discuss possible implications of our results with regards to the rehabilitation of CI users, including language acquisition and school performance. We also discuss data logging and its potential for clinical practice and further research.

In the future, CI data logging can help to better understand the connection between users' environments and their rehabilitation after cochlear implantation. Data logging can also open new possibilities to support the rehabilitation of people with CI.

Measuring by marking; the multimedia annotation tool ELAN

Han Sloetjes, Olaf Seibert

Max Planck Institute for Psycholinguistics
P.O. Box 310, 6500 AH Nijmegen, The Netherlands
Han.Sloetjes@mpi.nl, Olaf.Seibert@mpi.nl

Introduction

ELAN is a multimedia annotation tool developed by "The Language Archive" [1, 2], a unit of the Max Planck Institute for Psycholinguistics. It is applied in a variety of research areas in which audio and/or video recordings are the basis for qualitative and/or quantitative analysis of different modalities of communication. This paper aims at presenting a general overview of the tool with an emphasis on the most recent developments.

General overview of the tool

ELAN is a tool for manually adding textual annotations to audio and/or video recordings. The annotations are stored separately from the source media in XML files. The data model is tier based, allowing for multi-level, multi-participant annotation of time based media. A tier, in this context, is a kind of layer object, a container for grouping annotations that share the same topic or target and the same coding scheme. For example, one tier could contain annotations transcribing the speech of participant AA, another tier could record the left-hand gestures of participant BB etc. The user defines the tiers and determines what to use each for, which codes or categories to apply to each of them. The basic form of an annotation consists of a start time and end time, marking a specific segment of the media stream, possibly enriched by a text string. Annotations can also be linked to other annotations and be part of a hierarchically structured group of annotations. By marking segments of interest and storing them as annotations the user produces datasets that represent a measurement of the scored behavior. In many cases these measurements are the basis for further qualitative or quantitative analysis performed by other tools.

It is possible to link multiple video and multiple audio streams (e.g. in the case of recordings with multiple cameras and microphones) and the user interface allows to specify which videos and which sound channels to view and hear at any given time. ELAN is available for Windows, Mac OS X and Linux, is mostly written in the Java programming language, is free for anyone to use and is open source.

Linguistic Annotator

The "LA" in the ELAN acronym (originally: **E**udico **L**inguistic **AN**notator, where Eudico was the name of a project) indicate that the tool initially was conceived as a tool for annotating verbal human communication. And indeed an important part of the user base consists of (field) linguists. Considerable funding contributions came from a big project on endangered languages, DoBeS [3]. Within this project more than 60 teams were funded to document a language that is in danger of becoming extinct. The efforts of these teams were supported and facilitated by the development of software for archiving language data, producing metadata descriptions and for transcription of primary data. The aim was to produce transcription and translation annotations for most of the recordings and more elaborate morpho-syntactic glossing annotations for a smaller portion of the data. Nevertheless, most of the functionality of ELAN and the design of the data format are not particularly linguistic by nature but rather generic. To illustrate this, it is only to the present day that serious efforts are made to integrate a lexicon component (which adds linguistic functionality). That's how almost from the beginning the tool was also used for non-verbal human communication, with strong user communities in sign language research and in gesture and multimodal interaction research. Apart from that, but to a lesser extent, the application is also used for coding or scoring animal behavior.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

What these user groups have in common in terms of required functionality is segmenting and labelling (although labelling is not always required, e.g. in some multimodal interaction projects and animal behavior scoring segmentation on multiple levels suffices). Almost always annotation takes place on multiple tiers (e.g. for multiple participants and multiple modalities (speech, gesture, facial expressions)) and in many cases the feature of creating tier hierarchies (parent-child relations between tiers) is harnessed. Linguists often use dependent tiers for translations and for morpho-syntactical analysis layers and some gesture researchers subdivide gesture units into smaller units on dependent tiers.

But there are also requirements that are specific to certain user groups. Many linguists also use tools like Toolbox or FLEx [4] for transcription and morpho-syntactic glossing and for building the lexicon and they rely on the import and export functions for these formats for interoperability. An exhaustive description of Toolbox and FLEx functionality and on how to transfer data between these tools and ELAN is given by [5]. It is fairly common in endangered languages projects to create subtitled videos as one of the outcomes, therefore the export to subtitled text is important for such projects (ELAN does not create hardcoded subtitled videos).

Gesture and multimodal interaction researchers seem more likely to benefit from tier-based operations, features that allow to combine annotations on two or more tiers to create new annotations, e.g. annotations from overlaps, or annotations by merging or subtracting. These functions are useful for combining already created, existing segmentations for the purpose of creating new annotations representing a specific desired (sub-) result.

Interrater agreement or reliability

In some projects it is important to assess the reliability of annotations created by different annotators (inter-annotator) and/or annotations created by the same annotator but at different times (intra-annotator). For a long time ELAN did not have built-in functionality to calculate agreement except for a simple algorithm for comparing annotations on two tiers without chance correction. For more elaborate assessment of agreement or reliability other tools were (and are) available and exporting as .csv text suffices to feed the data into these tools.

One of these tools is EasyDIAg [6] implemented in Matlab in the context of NEUROGES [7] (gesture) research. The algorithm is based on Cohen's kappa, extended with a routine to match annotations based on the amount of overlap (therefor the segmentation decisions are part of the agreement assessment). Recently this algorithm has been (re-)implemented in Java and included in ELAN, with some additional flexibility in pairing tiers to compare. There is no predefined prefix or suffix on the basis of which tiers are combined and tiers do not have to be in the same file in order to be compared.

Another tool is Staccato [8] which calculates a Degree of Organization for segmentations (ignoring the labels) while applying a Monte Carlo simulation for randomization. This tool was already implemented in Java and was added to ELAN as a separate library, following a collaboration within a CLARIN-D [9, 10] working group for multimodality.

Manual vs automatic annotation

ELAN is tool for manual annotation of recordings; the usual workflow is that the user inspects the media stream, identifies relevant segments (depending on the area of interest and research) and creates annotations using the mouse and/or the keyboard. The level of precision completely depends on the user and on the nature of the research project. In some types of research it is not necessary that the segments are frame-precise and an overshoot of a few hundred milliseconds at start and end boundary are acceptable. The segmentation task can then be performed in the Segmentation Mode, in which segments are created by one or two key strokes while the media is playing. In other projects it is crucial that segment boundaries are as exact as possible, frame precise. This requires close inspection with a lot of frame stepping back and forth and playing a selection repeatedly before deciding on the segment. Most likely this is done in Annotation Mode. Especially the latter approach consumes a lot of the annotator's time and is therefore expensive.

Yet, this is still the default workflow in ELAN. Although we have been involved in several projects that aimed at development and integration of tools for (semi-)automatic segmentation and/or labelling, successful application thereof in real world scenario's proved to be problematic. Some algorithms perform fairly well when applied to "clean", good quality data, but perform poorly on naturalistic lab or field data.

We have been involved in two projects concerning automatic audio and video recognition, AVATecH and AUVIS [11], both in collabaration with two Fraunhofer institutes. A first version was included in 2010 and within the AUVIS project the algorithms were improved and the user interface further streamlined. These technologies can not replace manual creation of annotations yet, but can be applied assistively, in a scenario in which the automatic recognizer creates a segmentation which can then be manually corrected.

In the context of the big CLARIN project ELAN has been extended with client functionality for the text-processing tool-chaining framework WebLicht. Text can be uploaded to that webservice and the user can select one of the available tools for e.g. morpho-syntactic parsing and tagging of the text. The returned data are converted back to ELAN's data format and added as new tiers. This facility is mainly of interest for users who are working on (communication in) languages for which such taggers exist (so-called major or bigger languages).

Sharing Comments with team members

ELAN does not support true collaborative annotation, where multiple users simultaneously work on the same file. But to improve collaboration of team members who work on the same transcripts, either successively or in different files that are later merged, a commentary framework was implemented. A Comment is text that the user links to a segment of the media and possibly to a tier. Comments can be used as notes, remarks or questions concerning a fragment for later or for discussion with colleagues. This is quite similar to the way comments in word processors are applied. There are several ways in which comments can be shared: via email, via a file sharing (cloud) service (such as Dropbox) or via the DASISH Web Annotator (DWAN) framework [12, 13]. Comments can search and filtered based on one or several of their properties.

Export to Theme format

One of the more recently added export formats is that of Theme [14] data files. Theme is an application for detection and analysis of hidden patterns in time (so called t-patterns) in behavioral data. Annotations of (a selection of) tiers can be stored in a set of files that can be imported into a Theme project. A Theme project requires at least two text files, a Category or Variable-Value-Table File (.vvt, a file containing a listing of the coding Classes and the Items of each class) and one or more Raw Data files. The VVT is similar to the idea of Controlled Vocabularies and their entries in ELAN; when exporting the user can choose to export the entire CV's or just the values that were actually applied to annotations. The data text files, in which each line represents a record of a time-stamped event, contain the annotation data. Time-alignable annotations in ELAN have a duration (ELAN does not support single point annotations), therefore each annotation occurs twice in the output, once as a start ('b') event and once as an end ('e') event. The notion of Actor in Theme corresponds to either the tier name or the participant label of a tier (if it is there). This is a multiple file export, facilitating transfer of the data of a (sub-) corpus to Theme in one action.

Conclusions

ELAN is a multimodal annotation tool that is used in many behavioral research projects. It offers many functions that are indispensable to measure behavior on the basis of media recordings of (human) interaction. Despite participation in projects that aim at a shift towards automatic recognition of segments of interest, ELAN is still primarily an application in which the user has to mark the segments of interest manually. The user also has to determine the code or category for each segment, if any, resulting in a data set that represents a measurement of behavior. Manual annotation is a time consuming and expensive procedure, therefore it is inevitable that there will be more initiatives to improve the algorithms for automatic recognition. Both in the case of manual and automatic annotation the quality of the results has to be assessed therefore the new features for calculating inter-rater reliability are a useful extension of the program. The export to the Theme data format improved the level of interoperability with other tools that are relevant to the field. Some features that are in preparation are a lexicon component, cross-referencing between annotations and simplified user interfaces that expose only a limited set of functions, necessary for specific tasks. A simplified interface with reduced functionality will hopefully make the learning curve less steep so that it will become easier to involve more people in annotation tasks.

References

1. ELAN annotation tool, <http://tla.mpi.nl/tools/tla-tools/elan/>. Accessed 27 January 2016.
2. The Language Archive, <http://tla.mpi.nl/>. Accessed 27 January 2016.
3. DoBeS, Dokumentation bedrohter Sprachen, <http://dobes.mpi.nl/>. Accessed 27 January 2016.
4. FLEx, FieldWorks Language Explorer, <http://www.sil.org/resources/software_fonts/flex>. Accessed 20 January 2016.
5. Pennington, R. (2014). Producing time-aligned interlinear texts: Towards a SayMore–FLEx–ELAN workflow. <http://www.academia.edu>. Accessed 25 March 2016.
6. Holle, H., Rein, R. (2014). EasyDIAg: A tool for easy determination of interrater agreement. *Behavior Research Methods, 46(3).*
7. Lausberg, H., Sloetjes, H. (2015). The revised NEUROGES-ELAN system: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behavior Research Methods, DOI: 10.3758/s13428-015-0622-z.*
8. Lücking, A., Ptock, S., Bergmann, K. (2011). Staccato: Segmentation Agreement Calculator. In *Proc. of the 9th International Gesture Workshop, May 25-27, 2011, Athens, Greece*
9. Common Language Resources and Technology Infrastructure, <http://clarin.eu/>. Accessed 20 January 2016.
10. CLARIN Germany, <http://www.clarin-d.de/de/>. Accessed 20 January 2016.
11. Audiovisuelles Data Mining in multimodalen Sprachdaten, <https://tla.mpi.nl/projects_info/auvis/>. Accessed 27 January 2016.
12. Data Service Infrastructure for the Social Sciences and Humanities, <http://dasish.eu/>. Accessed 27 January 2016.
13. Lenkiewicz, P., Shkaravska, O., Goosen, T., Broeder, D., Windhouwer, M., Roth, S., Olsson, O. (2014). The DWAN Framework: Application of a Web Annotation Framework for the General Humanities to the Domain of Language Resources. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), May 26-31, 2014, Reykjavik, Iceland*
14. Theme, T-Pattern Analysis, <http://patternvision.com/>. Accessed 20 January 2016.

Multi-Brain BCI Games: Where to Go from Here?

Anton Nijholt

University of Twente, the Netherlands & Imagineering Institute, Iskandar, Malaysia, anton@imagineeringinstitute.org

Abstract

In this extended abstract we look at some issues associated with multi-brain computing. They concern the position of this field of research in Brain-Computer Interface (BCI) research in general, as seen in various BCI roadmaps. The issues concern various kinds of competition and collaboration in multi-brain computing, observations on social interaction and multi-brain computing, and observations on multimodal interaction and multi-brain computing. Illustrations of these issues will be taken from BCI game examples. In addition we introduce this BCI field of research with some historical notes on early artistic applications.

Introduction

Clinical research on brain-computer interaction has focused on patients and assistive technology for disabled users. How can we provide an ALS (amyotrophic lateral sclerosis) patient with the possibility to communicate with the outside world using his or her brains only? Or, how can an artificial limb or prosthetic device be controlled by thoughts only. In the early years of brain-computer interface research many other applications were considered. Artists used EEG (electroencephalography) devices to audify and visualize brain activity and gave it a role in real-time performances. Rather than having just one person's brain activity measured and used, they also thought of, designed, and implemented artistic applications where brain activity of two users was required in order to produce a desired result. An example is a situation where two persons have to synchronize their brain activity with the help of visual feedback in order to produce an aesthetically pleasing visualization or a command to activate a device. But it is as well possible that two persons explore an audio-visual landscape and discover how it can be changed by modulating their brain activity, playing with sounds and visualizations, discovering effects of joint actions and discovering effects of contrasting actions.

These activities took place in the late 1960s and early 1970s [1]. At that time there were already papers on EEG recording of intended movement ('Bereitschaftspotential' [2], intended modulation of alpha activity [3] and externally evoked brain activity [4]. Vidal mentioned several BCI controlled applications, including the navigation of a spaceship. In the years that followed interest in BCI for artistic applications decreased. What else could be added to what was already done? And, EEG detecting devices were costly and not that easily available. In the decades that followed we can see the further recognition in neuroscience and BCI research of various BCI paradigms that are now well-known and systematically used in BCI applications: event-related potentials (ERPs), evoked potentials (visual, auditory, touch, smell, taste) and intended modulation of brain activity by a user (trying to relax, pay attention to, focus on or perform a mental task, imagine a movement).

Although there are some examples after the 1970s of artistic use of BCI and having multiple users involved, only in the 21st century we see again a growing interest. Obviously, this was stimulated by the introduction of cheap BCI devices and accompanying software that produces (black box) results of detecting and interpreting brain activity, whether it is about affective state or intended command detection. Multi-brain computing became an issue in BCI game research, in neuro-marketing, group decision making, and evaluation of group performance. In this extended abstract we look at some issues associated with multi-brain computing. They concern the position of this field of research in BCI research in general, as seen in various BCI roadmaps. But more importantly, they concern observations on various kinds of competing and collaborating in multi-brain computing, on social interaction and multi-brain computing, and multimodal interaction and multi-brain computing. Illustrations of these issues will be taken from (BCI) game examples.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

Multi-Brain BCI: Applications

Slightly adapted from [5], we have the following applications of research on multi-brain computing: (1) Joint decision making in environments requiring high accuracy and/or rapid reactions or feedback, (2) Joint/shared control and movement planning of vehicles or robots, (3) Assess team performance, stress-aware task allocation, rearrange tasks, (4) Characterization of group emotions, preferences, appreciations. (5) Social interaction research (two or more people), (6) Arts, entertainment, games. In [6] we gave an arts and games biased survey of these applications. It is interesting to see how road maps for BCI research look at such multi-brain BCI applications. They are not or hardly mentioned, which is understandable since researchers who represent the BCI field look at clinical applications only and focus on solutions to problems of individuals (patients) [7]. Neuro-marketing is one of the fields that is sometimes mentioned, but clearly, this is not about (real-time) interaction, collaboration or competing.

Apart from multi-brain games that now are being developed [6] there is much interest by artists, often musicians, in having one or more persons' brain activity take part in creating or adapting a piece of life media art. This can be done in a performance that requires interaction with the artist/performer who is coordinating the joint life performance [8]. It is also possible that the artist delivers an interactive piece of art and it is up to the public to discover how to interact with it and possibly control the created interactive audio-visual (virtual or physical) environment [9]. Measuring audience responses with BCI is less well known than with more traditional physiological sensors (heart rate and skin conductivity sensors), but with the more recently developed EEG wireless headsets with dry electrodes this will certainly change. In games it is usually not about multiple players discovering interesting and entertaining properties of an environment. They are more aimed at competition (with one or more other players), collaboration (with one or more other players) or social interaction (with one or more other players) that is not or less task oriented than we see in interactions related to task or game collaboration and task or game competition.

Cooperation and Competition Using BCI

Whether it is about cooperation or competition, we think it is useful to distinguish between the following situations. Our examples address entertainment game interactions. We think these issues are important in non-game interactions as well, whether they are cooperating, competing or social.

First we distinguish between comparative and interactive games. In a comparative game there is no interference between the players. The performance of one player has no impact on the other player. When the game is played together or in the presence of an audience there may be social aspects that have impact. Consider for example playing with a pinball machine using imagery movement for pin control [10]. There is turn taking between players, but apart from score announcements and issues such as engagement and frustration, after each turn the game situation is neutral again.

This is different from turn taking in the Connect Four game. Players have a vertical grid of 6x7 positions and, taking turns, they can drop a coin in a chosen column. The game end when a player has four of his coins connected, either vertically, horizontally or diagonally. In [11] an event-related potential (P300) is used to let players choose a column. Once a coin is dropped the game state has changed, the opponent faces a new situation. This can be compared with playing chess and have your pieces moved by P300 or imagery movement [12]. The outcome of the Connect Four or chess game is not dependent on your BCI skills (unless we introduce time constraints).

BCI skills play an essential role in the game of Pong. There is turn taking, as in a tennis game, but it is turn taking with a time constraint (determined by the game, that is, the time it takes for the ball to move from one player to the other). Moreover, while your opponent is taking his turn, you can already prepare and in fact start your turn by predicting the ball's position when it returns from your opponent. There is a continuous change of the game state, but no control interference. Each player has its own bat to move. In [13] Pong players use motor imagery to move the bats.

Finally we mention games where there is BCI control interference and continuous change of the game state. A simple example is the BrainBall game where we have two competing players and BCI control through relaxation [14] is about the control of the ball and have it moved in the direction of your opponent. Other examples are mentioned in [6]. We think that especially this category of games, whether competitive, cooperative, or both, should be called BCI games. The outcome of the game is determined by the BCI skills of the participants.

Multi-Brain, Multi-Modal and Social Interaction

There are a three more observations we want to make when considering the future of BCI games (and more generally, the use of BCI in competitive and cooperative situations). Firstly there is the issue of fusion of information (brain activity) to be detected and interpreted from multiple players. It is an issue nicely discussed in [15], distinguishing between fusion at signal, feature and decision or application level. Obviously, possibilities are also dependent on whether players use the same BCI paradigm. In multimodal interaction research these issues have been investigated and in future multi-brain research we can profit from the knowledge that has been obtained there [16]. The second issue we want to mention is the impact of the cognitive or affective state of the participants on the issuing of BCI commands. The affective state of a user influences intended brain activity. In [17] two scenarios are mentioned: (1) accounting for the user's emotional state to adapt the algorithms that identify the user's intent in the ongoing EEG signals, and (2) purposefully eliciting emotion to enhance EEG features that are relevant for BCI. In a competitive situation we can also think of a scenario where we purposefully elicit emotion to weaken EEG features. Whatever scenario, we need to consider brain processes that concurrently occur. The final issue we want to mention is current neuroscience research on social interaction [18,19]. This research aims at the study of brain patterns that emerge during social interaction. This interaction can, for example, be a conversation, joint music making, or more generally, having a task that needs cooperation and joint attention. Again, as in the previous remark on concurrent occurring brain processes, here we have concurrent and often synchronized brain processes related to the interaction that takes place and that provide a context of interpreting intended and evoked modulations of brain waves.

Conclusions

In this short paper we had a look at issues that can play a role in future BCI games. They mainly concern how to deal with various concurrent brain processes in the players' brains. How can they enhance the players intentions during cooperation or how can they be employed to weaken players' intentions during competition? It is clear that the challenging issues addressed here are not yet at the foreground of current BCI and neuro-social interaction research. A longer version of this paper will appear in [20].

References

1. Rosenboom, D. (ed.) (1976). Biofeedback and the Arts: results of early experiments. A.R.C. Publications, Vancouver.
2. Kornhuber, H.H., Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Arch* **284,** 1–17.
3. Kamiya, J. (1968). Conscious control of brain waves. *Psychology Today* **1**(11), 56-60.
4. Vidal, J. (1973). Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, **2**, 157-180.
5. Stoica, A. (2012). MultiMind: Multi-Brain Signal Fusion to Exceed the Power of a Single Brain. Proceedings Third International Conference on Emerging Security Technologies, 94-98.
6. Nijholt, A. (2015). Competing and Collaborating Brains: Multi-Brain Computer Interfacing. Chapter 12: Brain-Computer Interfaces: Current trends and Applications. A.E. Hassanieu & A.T. Azar (Eds.), Springer,

Switzerland, 313-335.
7. Brunner, C., Birbaumer, N., Blankertz, B., Guger, C., Kübler, A., Mattia, D., Millán, J. del R., Miralles, F., Nijholt, A., Opisso, E., Ramsey, N., Salomon, P., Müller-Putz, G.R. (2015). BNCI Horizon 2020: Towards a Roadmap for the BCI Community. *Brain-Computer Interfaces* 2(1), 1-10.
8. Zioga, P., Chapman, P., Ma, M., Pollick, F. (2014). A Wireless Future: performance art, interaction and the brain-computer interfaces. Proceedings ICLI 2014 - Interface: Intern. Conf. on Live Interfaces, 220-230.
9. Wadeson, A., Nijholt, A., Nam, C.S. (2015). Artistic Brain-Computer Interfaces: Current State-of-Art of Control Mechanisms. *Brain-Computer Interfaces* **2**(2-3), 70-75.
10. Tangermann, M.W., Krauledat, M., Grzeska, K., Sagebaum, M., Vidaurre, C., Blankertz, B., Muller, K.R. (2008). Playing Pinball with non-invasive BCI. In: Proceedings Advances in Neural Information Processing Systems 21 (NIPS 2008), D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), Curran Associates, Inc., 1641-1648.
11. Maby, E., Perrin, M., Bertrand, O., Sanchez, G., Mattout, J. (2012). BCI Could Make Old Two-Player Games Even More Fun: A Proof of Concept with “Connect Four”. *Advances in Human-Computer Interaction*, Article ID 124728.
12. Maruthappan, N., Iyengar, N., Sudip Patel, P. (2011). Brain Chess – Playing Chess using Brain Computer Interface. In: ICBMG 2011 workshop, *IPCSIT* **20**, 183-191. IACSIT Press, Singapore.
13. Krepki, R., Blankertz, B., Curio, G., Müller, K.R. (2007). The Berlin brain-computer interface (BBCI)—towards a new communication channel for online control in gaming applications. *Multimed Tools Appl.* **33**(1), 73–90 (2007)
14. Hjelm, S.I., Browall, C. (2000). Brainball: using brain activity for cool competition. In: Proceedings of the First Nordic Conference on Computer-Human Interaction (NordiCHI 2000), Stockholm, Sweden.
15. Wang, Y., Jung, T.-P. (2011). A Collaborative Brain-Computer Interface for Improving Human Performance. *PLoS ONE* **6**(5): e20422, 1-11.
16. Turk, M. (2014). Multimodal interaction: A review. *Pattern Recognition Letters* **36**, 189-195.
17. Garcia Molina, G., Tsoneva, T., Nijholt, A. (2009). Emotional Brain-Computer Interfaces. In: Proceedings 3rd International Conference on Affective Computing and Intelligent Interaction - ACII 2009, IEEE Computer Society Press, 138-146.
18. Babiloni, F., Astolfi, L. (2012). Social neuroscience and hyperscanning techniques: past, present and future. *Neurosci Biobehav Rev.* **44**, 76-93.
19. Mattout, J. (2012). Brain-computer interfaces: a neuroscience paradigm of social interaction? A matter of perspective. *Front. Hum. Neurosci.* **6**, Article ID 114.
20. Nijholt, A., Poel, M. (2016). Multi-Brain BCI: Characteristics and Social Interactions. In: 10th International Conference on Augmented Cognition, 17-22 July 2016, Toronto, Canada. Lecture Notes in Computer Science. Springer Verlag, Berlin, to appear.

Continuous Affect State Annotation Using a Joystick-Based User Interface: Exploratory Data Analysis

K. Sharma[1], C. Castellini[1] and Egon L. van den Broek[2]

[1]Robotics and Mechatronics Center, German Aerospace Center, Oberpfaffenhofen, Germany
[2]Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
karan.sharma@dlr.de and vandenbroek@acm.org

Abstract

The DLR (German Aerospace Center) aims to assess user's affective state in motion simulators. To facilitate this goal, a joystick-based user interface was used to gather reports on user's emotions. This user interface allowed continuous annotations, while video clips were watched. In parallel, several physiological parameters (e.g., electrodermal activity, heart and respiration rate) were acquired to record affective responses. An exploratory data analysis of the users' ratings (incl. several visualizations) that unveils several interesting data patterns is presented.

Introduction

The work presented here is a continuation of the research summarized in our submission to *Measuring Behavior 2014* [1]. Hence, a short re-introduction to this preceding work is presented in this section. The main focus of the aforementioned submission was to introduce the Data Acquisition (DAQ) and the video-playback system that we have developed for undertaking tests involving 'affect elicitation from videos' conditions. The two main aspects of this system are as follows:

1. **The Annotation User Interface (UI)**: features the annotation interface embedded in the video playback screen. More details on the underlying concepts and design of the UI can be found in [1].

2. **The Data Acquisition System:** simultaneously acquires the participants' physiological parameters (e.g., electrodermal activity, heart and respiration rate) [1,2,4,7-10,14,18] and self-reported affect state through the joystick. For the same, the participants were instructed to position the joystick on a 2-D plane such that the x and y position of the joystick indicates the experienced valence and arousal levels, respectively (cf. [15]).

Traditionally, experimental studies investigating participants' affective response predominantly tend to use Likert scale based self-reporting techniques, wherein the participants report their affect states on questionnaires post-stimuli [3,4,10]. Some advantages of our joystick based method over these techniques are:

1. A video is a dynamic stimulus [2,4,11,20], therefore continuous self-reporting by the participants allows us to link their responses to the events in the video and also to investigate how their affect state evolved during the course of the video, thereby providing more insight into the data [5,9,12,17].

2. A joystick based annotation system is intuitive to use (cf. [19]) and allows the participants to report their affect state at the same time as it is elicited [17].

3. As the participants report their perceived valence and arousal levels instead of discrete emotional states, the UI can also account for mixed emotional experiences [6,13-17].

A preliminary data analysis of the annotation data (i.e., the joystick ratings) recorded during the experiment was presented in an earlier submission [1]. In this current work, we extend the previous preliminary analysis by undertaking an exploratory data analysis (EDA) of the same annotation data. To perform the EDA, the data is first pre-processed and thereafter summarized. Then, exploratory graphs of this summarized data are presented.

These graphs provide insights into the underlying main characteristics of the data and are helpful in determining the appropriate methods to be used for model fitting and hypothesis testing.

Data Pre-processing

Before EDA can be performed on the annotation data, the data needs to be pre-processed and labelled [3,13]. The pre-processing step involves:

1. Using the *ffprobe* tool from the FFmpeg multimedia framework to determine the exact duration of the video sequences (unique for every participant) as well as the individual videos.

2. Using the duration information extracted in the last step, video-labels are added to the annotation data that contains joystick timing and co-ordinate position data. The corresponding video-label information needs to be added to this data for further data analysis. The resulting labelled joystick data for a single subject is shown in Figure 1.

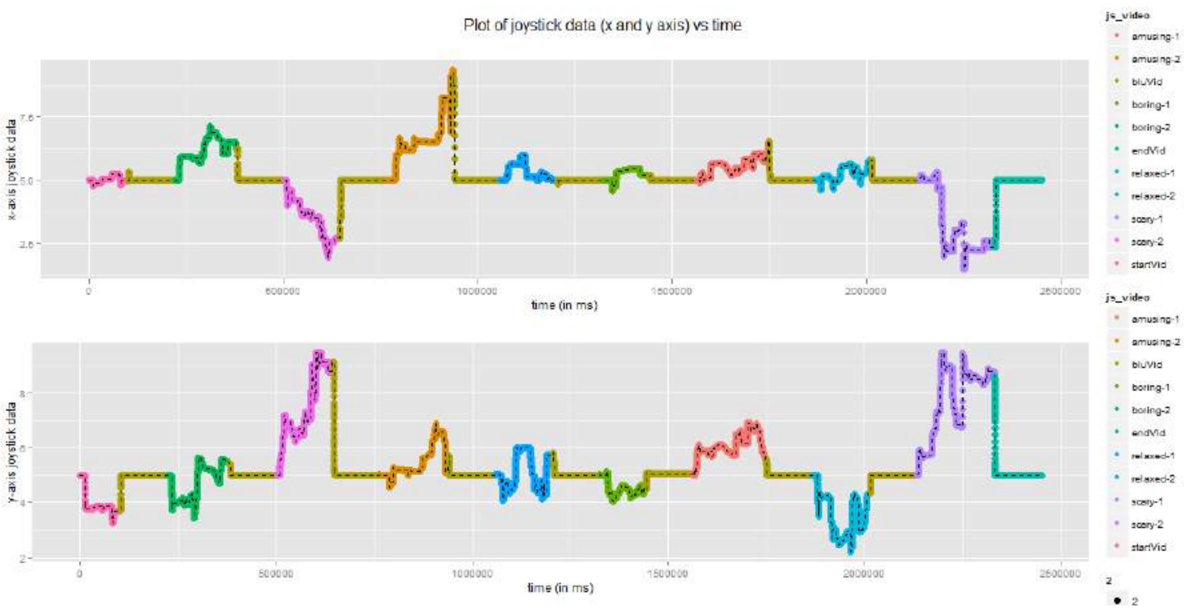


Figure 1. Plot of the joystick data-points (connected by a dashed black line) for one subject.

Data Analysis Levels

It can be seen in Figure 1, that the labelled joystick data is comprised of x and y position time series. This data can be analysed at two levels [14,17,20]:

1. **Mean ratings per video**: the *mean* valence and arousal ratings (x and y-axis joystick data, respectively) by a participant for every video are computed by calculating the arithmetic mean of the x and y positions across the complete time series (as seen in Figure 1). As there are 30 participants and 8 videos of interest in the study; this computation results in 30 mean ratings (i.e. 30 values of $\bar{x}$ and $\bar{y}$ each) for each video, where each participant provides mean ratings for 8 videos.

2. **Time series data:** continuous self-reporting through the joystick results in valence and arousal time series. By analysing these time series, a participant's response can be precisely co-related to the events in the video, and the evolution of user affect states during a video can be better analysed.

A detailed EDA of the time-series data is beyond the scope of this publication and hence will not be presented in the following sections. In the remainder of this publication, an EDA of the *mean ratings per video* data is presented.

Exploratory Data Analysis (EDA)

Figure 2 contains multiple scatter plots of the mean ratings for each of the 8 videos. In these plots, each black dot represents a mean valence and arousal rating ($\bar{x}$ and $\bar{y}$, respectively) from one subject for that video. Therefore, all the 8 scatter plots contain 30 data points each.

The scatter plots show the spread of the mean ratings for different types of videos. The blue line in these plots represents a simple linear regression line that has been fit to the given data. The shaded grey areas around the blue lines depict the 95% confidence region of the regression fit. The blue regression lines in these scatter plots show the relationship between valence and arousal ratings for different videos.

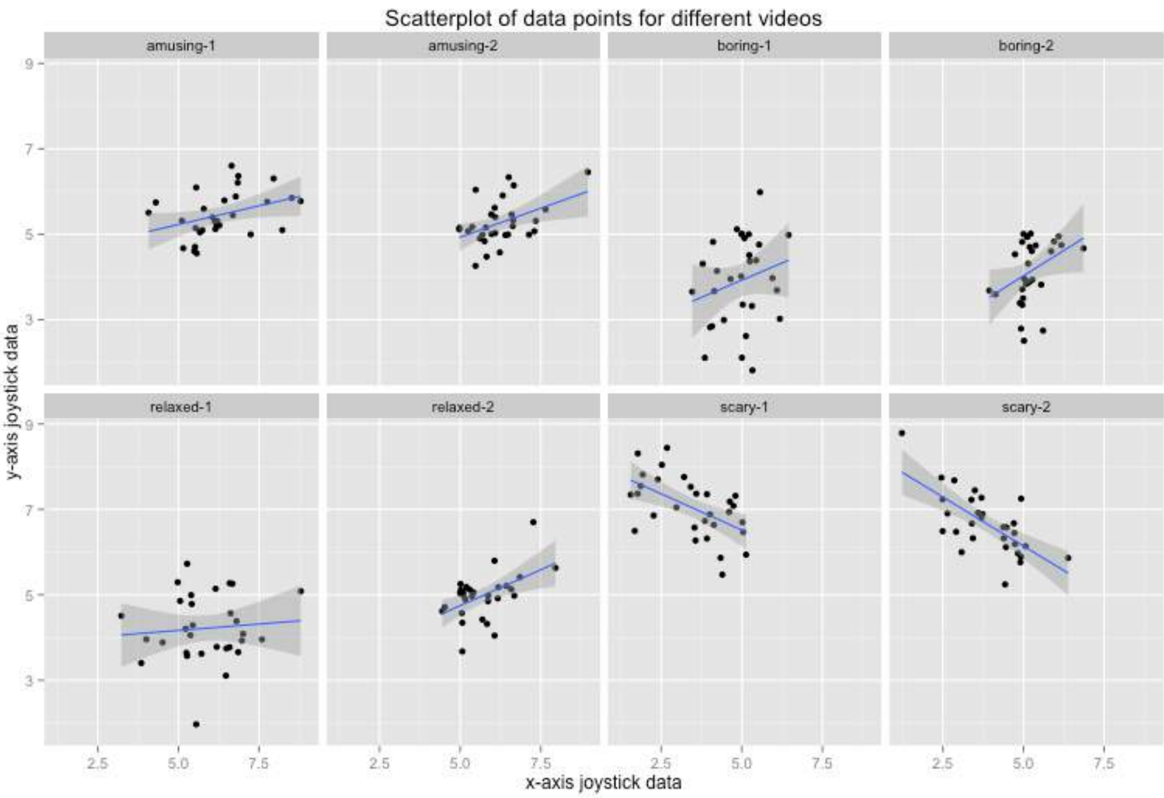


Figure 2. Scatter plot with regression lines (in blue) for different types of videos.

The box plots for the mean ratings are shown in Figure 3. This figure contains two box plots: the left and the right sub-figures show the box plots for the mean valence ($\bar{x}$) and arousal ($\bar{y}$) ratings, respectively.

In these box plots, solid black dots signify outliers; solid coloured boxes signify the Interquartile Range (*IQR*) of the data; thin black lines/whiskers signify the range Q1 - 1.5 *IQR* -- Q3 + 1.5 *IQR*; horizontal lines in the box signify the median of the mean values; and the star points signify the mean of mean values (cf. [14,18]).

Box plots are useful in visualising the spread (variance) of data in different categories (videos in our case) and facilitate a visual comparison of the data distributions across different categories.

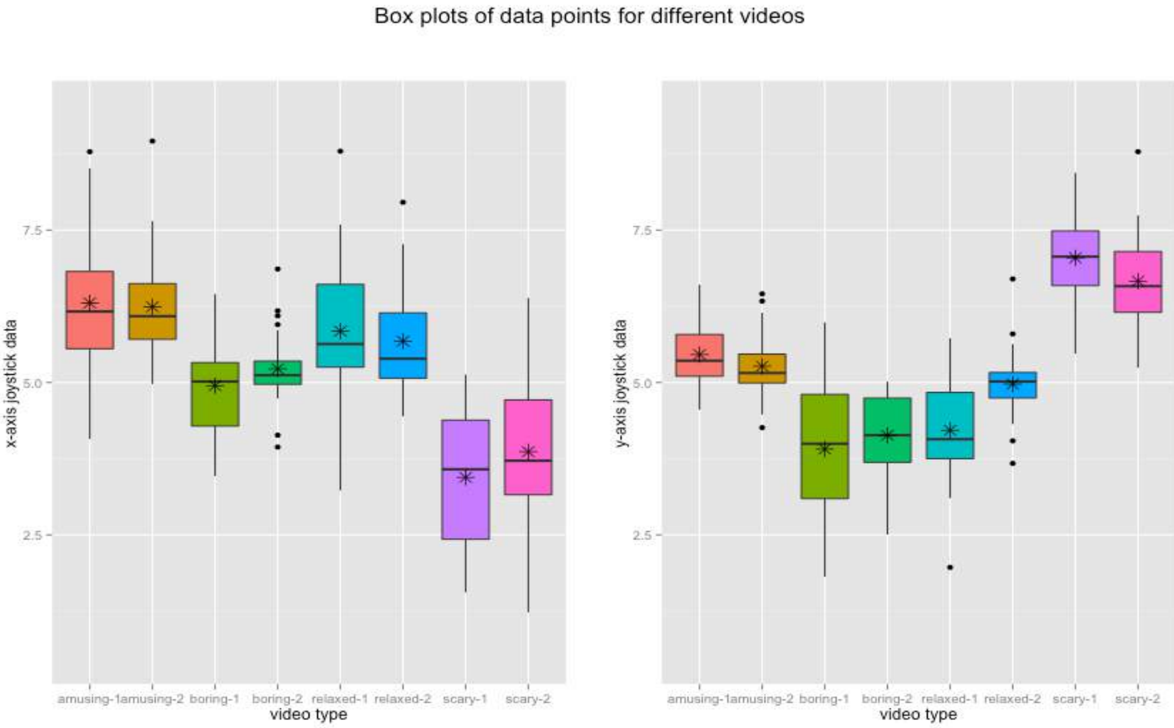


Figure 3. Box plots for the mean valence (x-axis) and arousal (y-axis) ratings.

Results

The scatter plots in Figure 3 show the location and spread of the mean ratings across different videos. For example, the mean ratings for scary-1 and 2 videos generally tend to be in the second quadrant; whereas, the ratings for amusing-1 and 2 videos tend to be in the first quadrant. The relationship between the mean ratings also differs across videos: for boring-2 video the variation of data in the *x*-axis (valence ratings) is less than the variation in the y-axis (arousal ratings); whereas, the variation of data in both axes is generally the same for the scary-1 video.

The distribution of the ratings data is better illustrated by the box plots. The data is said to be normally distributed if among others, the median is equal to the arithmetic mean. For example, for the scary-1 video, the mean and the median of the arousal (*y*-axis joystick data) ratings overlap with each other and both of them are approximately at the centre of the *IQR*. Based on these observations, we can state that arousal ratings for scary-1 video should be approximately normally distributed. Similarly, boxes that are larger in size than others have a larger *IQR*, which in turn implies that the distribution of the data is more spread out [14].

The skewness of a distribution can also be determined using the box plots (cf. [14,17,20]). Generally, if the median is not in the centre of the *IQR* box, the data is either negatively or positively skewed. For example, the median of the valence (*x*-axis joystick data) ratings for boring-1 video is not at the centre of the *IQR*, but rather shifted towards the top of the *IQR* box. This implies that the distribution is skewed to the left (negative skew).

The outliers (black dots in the box plots) in the data skew the mean in their direction; therefore, for distributions with large number of outliers the mean is skewed in one direction and hence is different from the median of the data e.g. arousal (*y*-axis joystick data) ratings for the amusing-2 video.

Conclusions

Given the multivariate nature of the presented dataset, a suitable model for hypothesis testing would be the Repeated Measures (RM) Multivariate Analysis of Variance (MANOVA) model. The EDA presented in this submission is an important initial step in determining if a chosen model is appropriate for the given dataset. For example, based on whether the data is normally distributed or not, either a parametric or a non-parametric analysis approach is chosen for the hypothesis tests.

To precisely determine which model is appropriate, thorough assumption testing must be undertaken. Nevertheless, the results of an EDA are essential building blocks that lead to comprehensive assumption testing. For example, one of the main assumptions for parametric MANOVA is that the correlation between any two dependent variables is the same in all groups of data. Using the scatter plots generated during the EDA, we can hypothesize that for the given dataset, this assumption might not be always fulfilled. Similarly, another assumption regarding the normality of data can be addressed through the box plots. The box plots presented here provide an initial indication regarding the normality or non-normality of the data. This initial indication can be then affirmed using specific tests.

Based on the EDA methods presented in this submission, an initial indication regarding the characteristics of the data was drawn. The EDA can be extended by including tests that check for specific assumptions related to models for hypothesis testing. These hypothesis tests will allow us to draw important insights and inferences about the joystick-based annotation system.

References

1. Antony, J., Sharma, K., van den Broek, E. L., Castellini, C., and Borst, C. (2014). Continuous affect state annotation using a joystick-based user interface. *Proceedings of Measuring Behavior 2014: 9th International Conference on Methods and Techniques in Behavioral Research* (Wageningen, The Netherlands), 268–271.
2. Bellmann, T., Heindl, J., Hellener, M., Kuchar, R., Sharma, K., Hirzinger, G. (2011). The DLR Robot Motion Simulator Part I: Design and Setup. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA 2011)* (Shanghai, China, May 9-13), p. 4694 –4701.
3. Bradley, M.M. and Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25(1), 49-59.
4. Janssen, J.H., Tacken, P., De Vries, G.-J., Van den Broek, E.L., Westerink, J.H.D.M., Haselager, P., IJsselsteijn, W.A. (2013). Machines outperform lay persons in recognizing emotions elicited by autobiographical recollection. *Human-Computer Interaction* 28(6), 479-517.
5. Janssen, J.H. Van den Broek, E.L., Westerink, J.H.D.M. (2009). Personalized affective music player. In J.F. Cohn, A. Nijholt, M. Pantic (Eds.), *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction (ACII2009)*, 472-477. Los Alamitos, CA, USA: IEEE Computer Society.
6. Lang, P.J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist* 50(5), 372-385.
7. Levenson, R.W. (1988). Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity. In H.L. Wagner (Ed.), *Social psychophysiology and emotion: Theory and clinical applications, Chapter 2*, 17-42. Chichester, West Sussex, Great Britain: John Wiley & Sons Ltd.
8. Liu, C., Rani, P., Sarkar, N. (2006). Human-robot interaction using affective cues. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)* (Hatfield, UK, September 6-8), p. 285–290.
9. Nagel, F., Kopiez, R., Grewe, O., Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods* 39(2), 283-290.
10. Picard, R.W. (1995). *Affective computing*. Technical Report 321, M.I.T Media Laboratory Perceptual Computing Section, Cambridge, MA, USA.
11. Rottenberg, J., Ray, R.D., Gross, J.J. (2007). Emotion elicitation using films. In J.A. Coan, J.J.B. Allen

(Eds.), *Handbook of emotion elicitation and assessment, Chapter 1*, 9-28. New York, NY, USA: Oxford University Press, Inc.

12. Ruef, A.M. and Levenson, R.W. (2007). Continuous measurement of emotion: The affect rating dial. In J.A. Coan, J.J.B. Allen (Eds.), *Handbook of emotion elicitation and assessment, Chapter 17*, p. 286-297. New York, NY, USA: Oxford University Press, Inc.
13. Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review* 110(1), 145-172.
14. Van den Broek, E.L. (2011). *Affective Signal Processing (ASP): Unravelling the mystery of emotions.* CTIT Ph.D.-thesis series No. 11-204, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.
15. Van den Broek, E.L. (2013). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing* 17(1), 53-67.
16. Van den Broek, E.L., Lisý, V., Janssen, J.H., Westerink, J.H.D.M. , Schut, M.H., Tuinenbreijer, K. (2010). Affective Man-Machine Interface: Unveiling human emotions through biosignals. In A. Fred, J. Filipe, H. Gamboa (Eds.), *Biomedical Engineering Systems and Technologies,* p. 21-47. Berlin/Heidelberg, Germany: Springer-Verlag.
17. Van den Broek, E.L., Westerink, J.H.D.M. (2009). Considerations for emotion-aware consumer products. *Applied Ergonomics* 40(6), 1055-1064.
18. Van Drunen, A., Van den Broek, E.L., Spink, A.J., Heffelaar, T. (2009). Exploring workload and attention measurements with uLog mouse data. *Behavior Research Methods 41(3)*, 868-875.
19. Vuurpijl, L., Schomaker, L., Van den Broek, E.L. (2002). Vind(x): Using the user through cooperative annotation. In A.D. Williams (Ed.), *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02)*, 221-226. Los Alamitos, CA, USA: IEEE Computer Society.
20. Westerink, J.H.D.M., Van den Broek, E.L., Schut, M.H., Van Herk, J., Tuinenbreijer, K. (2008). Computing emotion awareness through galvanic skin response and facial electromyography. In J.H.D.M. Westerink, M. Ouwerkerk, T. Overbeek, W.F. Pasveer, B. de Ruyter (Eds.), *Probing Experience: From academic research to commercial propositions*, p. 149-162. Dordrecht, The Netherlands: Springer Science Business Media B.V.

mPATHway: A Formative Research Process to Develop a Mobile-Application Based Intervention for Cardiovascular Disease

Orlaith Duff[1,2], Deirdre Walsh[1,2], David Monaghan[2], Catherine Woods[1], Kieran Moran[1] and Noel E. O'Connor[2]

[1] School of Health and Human Performance, Dublin City University, Dublin, Ireland.

Orlaith.duff2@mail.dcu.ie

[2] Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland.

Abstract

Cardiovascular disease (CVD) is the most common cause of mortality in Ireland, accounting for one third of all deaths and one in five premature deaths [1]. With the prevalence of CVD so high, not only in Ireland but across the world, cardiac rehabilitation has never been so important as a continuous process of care. The main purpose of cardiac rehabilitation is to prevent a further cardiac event and improve the person's quality of life. However, uptake of such programmes remains low. mHealth technologies may tackle some of the issues relating to poor uptake and low adherence, such as accessibility and affordability. This paper outlines the formative research process to develop a mobile-application for cardiovascular rehabilitation.

Introduction

Cardiovascular disease is the leading cause of premature death and disability in Europe, accounting for four million deaths per year and costing the EU economy almost €196 billion annually [2]. Exercise-based Cardiac Rehabilitation (CR) is used to reduce the impact of CVD and to promote healthy behaviours and active lifestyles for those with CVD [3]. Whereas, CR improves mortality and morbidity rates, adherence within these programmes is generally low [4]. Some of the more common issues identified with adherence to CR programmes relate to accessibility and parking at local hospitals, a dislike of group environments and work or domestic commitments [4].

Mobile health (mHealth) is an emerging area of healthcare and is defined as "*medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants and other wireless devices*" [5]. mHealth technologies have the ability to make healthcare more accessible, affordable and available to the public [6]. Although this is still a developing area of research, literature in the area of internet and mobile-based health interventions has found that such tools can be useful in supporting the self-management of chronic disease [7][8].

Within this paper we outline the development work within the mPATHway project. mPATHway is designed to utilize the expertise in a community-based medical exercise programme, called MedEx, and the Insight Centre for Data Analytics, to allow people with CVD to participate in an exercise-based rehabilitation programme remotely, through a specially designed Android App called MedFit. MedFit offers the potential to make exercise-based rehabilitation programmes more effective by making them more accessible, more personalised and more interactive, by providing real-time support and feedback for participants.

In order to develop an evidence-based effective mobile App for CVD embedded within health behaviour change theory, it is important to engage in a formative research process. Formative research involves key research activities including, a systematic review of current literature, usability testing, feasibility studies as well as extensive piloting and user validation. The key stages are outlined below and provide a best practice framework for App development within a health behaviour change and public health setting.

Formative Research Process

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

The formative research process is a critical step in the development of health behaviour change interventions and consists of 4 key stages; development, feasibility/piloting, evaluation and implementation [9].

The first stage of the research involves conducting a systematic review in order to identify the relevant, existing evidence base, as well as potential gaps in the literature [9] (see Figure 1.). Within mPATHway a crucial step of the development was to systematically examine previous research to assess mHealth and eHealth (electronic health) interventions to identify what behaviour change techniques are applied and to identify how effective they are in delivering meaningful long term behavioral change. The review entitled a 'Systematic review of the use of Behaviour Change Techniques (BCT's) in physical activity eHealth interventions for adults with cardiovascular disease' was thus registered in an online systematic review protocol database 'Prospero' as per best practice guidelines to aid transparency within the development process. This research will identify the most effective BCT's in eHealth interventions for CVD patients. Subsequently the results of this study will then be translated into technical requirements to inform the development and functionality of the application 'MedFit'. This means that both the technical aspects of the application (i.e., push notifications and the user interface) as well the tailored interactive content are all based within an evidence and theory-based framework of health-behaviour change. To date, preliminary searches have been carried out, piloting of the study selection process, as well as formal screening of the title and abstract of search results against the eligibility criteria.

In the next stage of the development process, the feasibility and acceptability of the prototype application will be tested in focus groups and subsequently in feasibility and pilot trials prior to the full scale trial. The focus group content will be developed and informed using models of technology acceptance. This stage of user validation through the focus groups is a crucial part of the user-led formative research and design process, with the purpose being to gain feedback on the first prototype of the mobile App. This feedback will be translated into feasible technical improvements through close collaboration with the technical design team, who will adapt and make modifications and upgrades to the app based on the patient feedback and comments from the focus group. This user validation will be an iterative process with subsequent rounds of focus groups, where the initial group are brought back to view the changes and additional features added to the App and provide additional feedback on these. This iterative design process with the end-user allows for the custom design and creation of a truly patient-centric home-based exercise-rehabilitation CVD platform.

Following this, the updated version of mPathway App will be evaluated through small-scale feasibility trial and pilot trials as per the Medical Research Council Guidance for the development and evaluation of complex behavioural change interventions [9]. A feasibility trial will be conducted in order to assess the intervention components and to access acceptability, likely rates of recruitment, retention of participants and to calculate appropriate sample sizes [9]. The feasibility trial will be conducted with collaborator hospital partners in the Dublin Region, with approximately 40 participants (intervention N=20, Control N=20) for a period of a minimum of 4 weeks to trial user engagement and adherence to using the app. Such research may identify any methods or protocols which need modification and also how changes might occur [10]. Following the feasibility trial medical practitioners will be approached for feedback in terms of likely rates of recruitment, retention of participants and any informal feedback they may have received from patients regarding MedFit. This will then inform the development of the pilot MedFit intervention, before a final full scale intervention can be carried out to evaluate the mobile application's effectiveness.

Conclusion

As mPATHway is currently an on-going and active research project, this research will focus on the phases of development and evaluation of a complex mHealth intervention. This entails the design process, first alpha-version of the App, focus group feedback, as well as a framework for early stage usability testing.

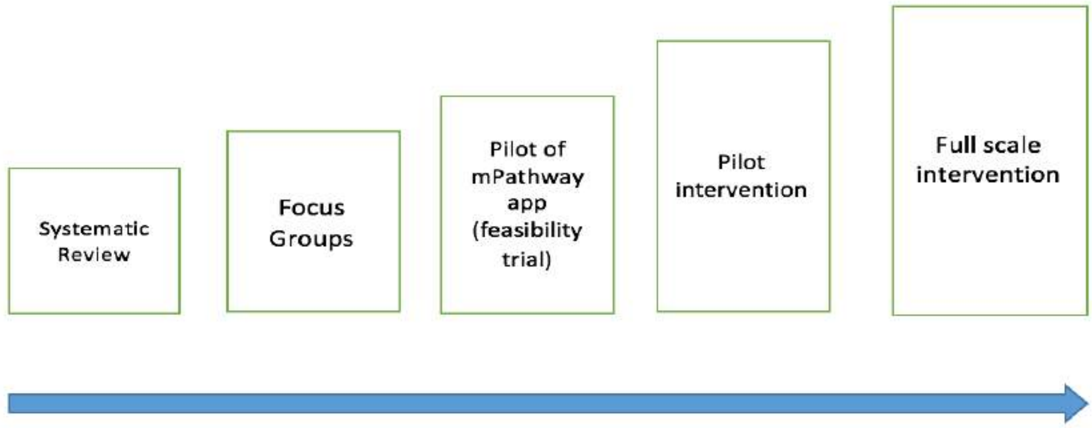


Figure 1. Stages of the formative research process

Acknowledgement

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289

References

1. Department of Health and Children. (2010). *Changing Cardiovascular Health. National Cardiovascular Health Policy 2010-2019.* Dublin: Government Publications.
2. Nichols M, Townsend N, Luengo-Fernandez R, Leal J, Gray A, Scarborough P, Rayner M (2012). *European Cardiovascular Disease Statistics 2012*. European Heart Network, Brussels, European Society of Cardiology, Sophia Antipolis
3. Balady, G., Ades, P., Comoss, P., Limacher, M., Pina, I., Southard, D., et al. 2000. Core components of cardiac rehabilitation/ secondary prevention programs: a statement for healthcare professionals from the American Heart Association and the American Association of Cardiovascular and Pulmonary Rehabilitation Writing Group. Circulation, 102: pp.1069-1073.
4. Dalal, H., Zawada, A., Jolly, K., Moxham, T., & Taylor, R. 2010. Home based versus centre based cardiac rehabilitation: Cochrane systematic review and meta-analysis. *British Medical Journa*l, pp.1-15.
5. World Health Organisation. 2011. mHealth: New horizons for health through mobile technologies. *Global Observatory for eHealth Series (Vol.3).* Geneva: WHO.
6. Akter, S., D'Ambra, J., & Ray, P. (2010). User perceived services quality of mHealth services in developing countries. *Proceedings of the Eighteen European Conference on conference on Information System.* Pretoria, South Africa: University of Pretoria.
7. Fanning, J., Mullen, S.P. and McAuley, E. 2012. Increasing physical activity with mobile devices: A meta-analysis. *Journal of Medical Internet Research,* 14(6), pp.61-71.
8. Omboni S, Gazzola T, Carabelli G, Parati G. Clinical usefulness and cost effectiveness of home blood pressure telemonitoring: meta-analysis of randomised controlled studies. *Journal of Hypertension,* 31(3), pp. 455-467.
9. Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ: British Medical Journal*, *337*, a1655. http://doi.org/10.1136/bmj.a1655
10. Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken S., Kaplan C., Squiers L., Fabrizio C., and Fernandez, M. (2009). How We Design Feasibility Studies. *American Journal of Preventive Medicine*, *36*(5), 452–457. http://doi.org/10.1016/j.amepre.2009.02.002

Dashboard visualisation of lifelog data for summarisation and pattern recognition to promote behavioural change

Aaron Duane, Cathal Gurrin and David Monaghan

Insight Centre for Data Analytics,
Dublin City University, Dublin 9, Ireland

Abstract

Lifelogging is a form of pervasive computing, which is capable of recording a catalogue of the totality of an individual's experiences. Lifelogging data can take many forms from daily step counts, sleep monitoring and heart rate data to location based information to video/image diaries from cameras embedded into glasses or clothing or worn. From the perspective of behavioural change, lifelog data is an invaluable resource that can contain many potential insights into a person's patterns and experiences. One major issue with Lifelogging is the sheer volume of user generated data and analysing this wealth of data in a coherent and efficient manner can become increasingly difficult as the datasets become larger and more varied. Through utilising data visualisation techniques in a dashboard interface, or similar, one can make high-level analyses and summarisations of this data much more achievable. This visually intuitive approach to disseminating rich datasets alongside automated and algorithmic machine-learning techniques can lead to valuable user insights and the potential to significantly promote behavioural change.

Introduction

Lifelogging is "a form of pervasive computing, which generates a unified digital record of the totality of an individual's experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive" [1]. The recording of this data has become an increasingly popular phenomenon in recent years due to the advent of personal sensors (e.g. accelerometers, heart rate monitors, global positioning systems) often conveniently embedded in a user's phone or wristband. A single lifelogger can accumulate thousands of images per day and when paired with their myriad other sensor data, the dataset, although potentially very valuable, can quickly become too large to conveniently manage. It is evident that one of the biggest challenges encountered when attempting to analyse this data is the sheer scope of it's size and variance.

Recent techniques to automatically extract actionable information from such image and sensor streams utilise machine learning algorithms and image recognition software, to scan an image and extract potentially relevant terms from a lexicon and assign them confidence values. These terms are referred to as 'lifelog concepts'. These concepts, along with the images and other sensor data, can form a faceted classification system that is ideally suited to data filtering and search. However, while this process is effective in further disseminating the richness of the data, the method is automated and therefore is always subject to varying degrees of accuracy.

Effective visualisation of data can alleviate many of the complications encountered when working with datasets of this nature. In this work, we present a dashboard interface designed to expose visually intuitive methods of exploring lifelog data. The data consists of physical activities, semantic locations and wearable camera images (about 1,500 per day). The interface itself relies on the automatically generated 'lifelog concepts' to both summarise and contextualise the data at any given time. Combined with the additional time and location metrics, we can isolate specific patterns and recurrences which can be further explored via additional visualisation techniques. The insights garnered via this method of analysis can provide knowledge to the individual, which can support effective behavioural change under a range of criteria.

System Description

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

The interface to visualise lifelog data is divided into three distinct sections. Each of the sections from left to right focuses on a more granular aspect of the lifelog data. The first and leftmost section of the dashboard summarises all of the algorithmically detected 'lifelog concepts' that are related to the data and ranks them in an ordered list from most common to least common (see fig. 1). The numbers next to each concept correspond to the amount of images where the term was detected. In this dataset each minute can only contain one image and each image contains up to fifty unique concepts. Each concept in this list can be toggled into an active state which will then highlight corresponding visualisations in the dashboard.

Figure 10. The Concept List

The second section, which is positioned in the centre of the dashboard, displays a high-level overview of each day of lifelog data. A day contains 1,440 individual segments representing each minute of that day (see fig. 2), from 00:00 to 23:59.

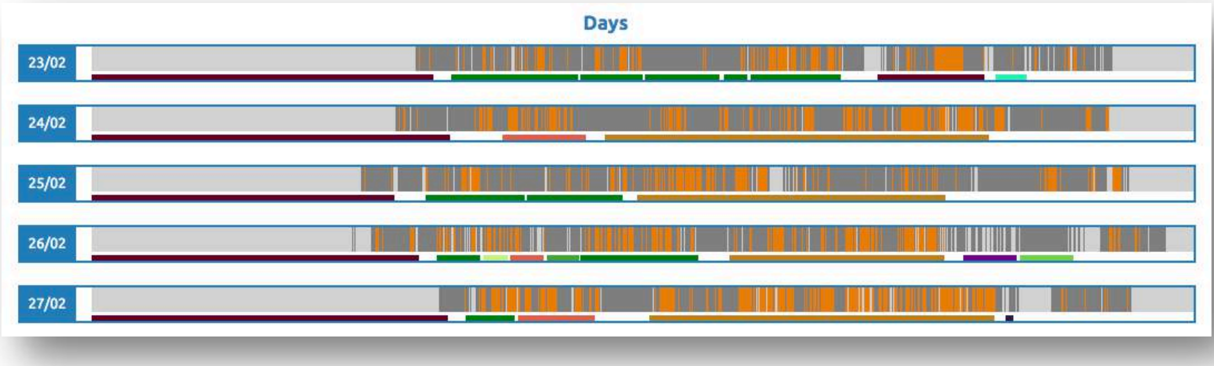


Figure 11. Every Day in 1,440 Segments.

In their default state each minute is represented by a single vertical line of grey. However, if an image was captured during that minute, it is conveyed to the viewer by a darker shade of grey. The minutes highlighted in orange correspond to the currently active concepts toggled by the user in the previous section. Selecting any concept from the ranked list immediately updates this section to indicate which minutes contain all currently toggled concepts.

Directly beneath each day is related visualisation element that indicates the detected location of the lifelog data at any given time. If the user hovers their cursor over a selected minute of interest, a tooltip is displayed conveying that current time and location (see fig. 3).

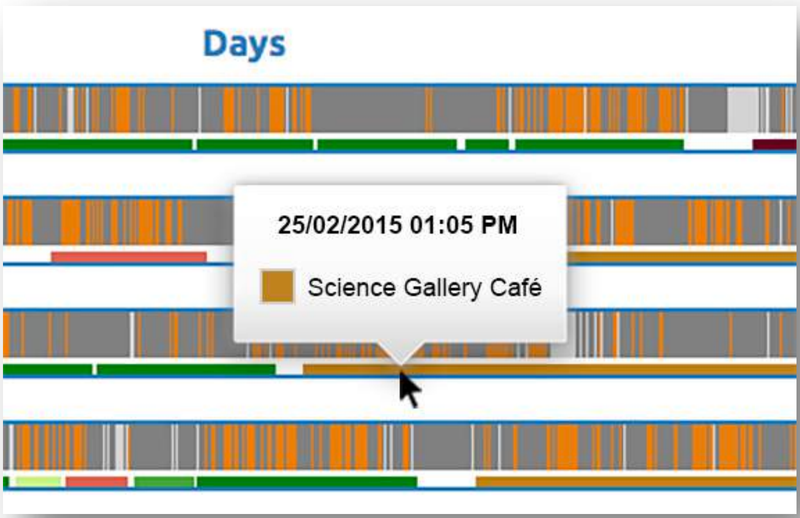


Figure 12. Illustrating the use of Additional Sensor Data

The third and rightmost section of the dashboard is interacted with by clicking on an individual minute the user wants to further analyse. Upon selection, the third (rightmost) section updates with any known metadata specific to that minute (see fig. 4).

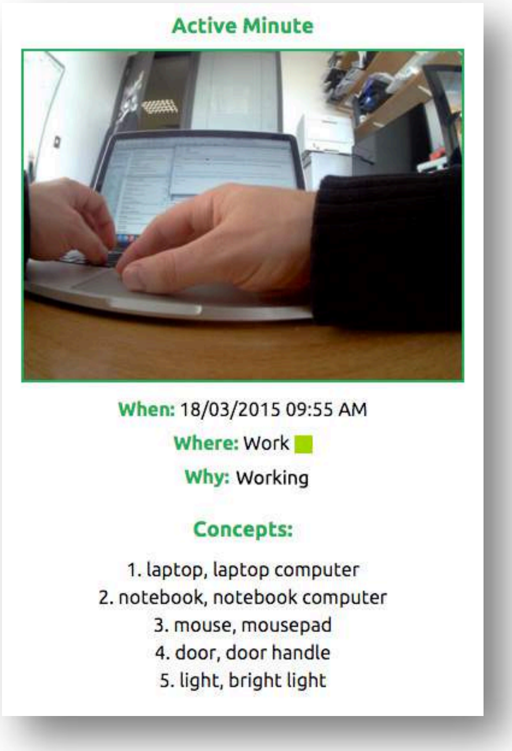


Figure 13. Metadata View of any Minute

If the minute contains an image, it is displayed to the user here at a slightly reduced resolution. This in turn can be clicked on if the full resolution needs to be examined by the user. Beneath the image is a list of all available metadata associated with the minute. In this particular dataset a minute can contain a time, location, an activity and the top five most confident lifelog concepts that were detected in the image.

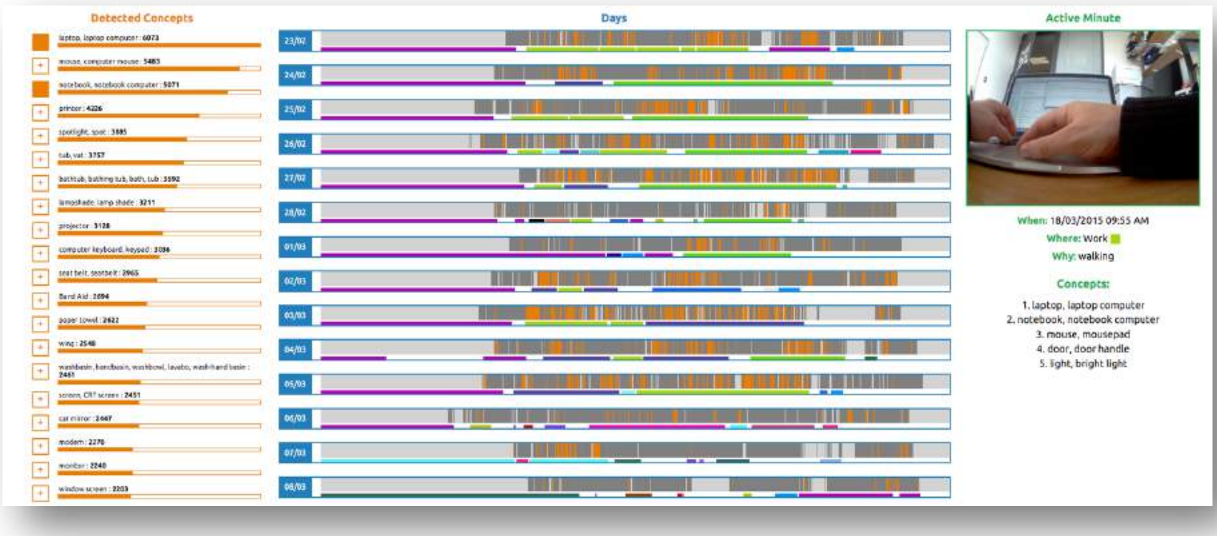


Figure 14. The Entire Interface View

Conclusions & Future Work

In this work we present an interface for visualising multi-sensor lifelog data. This interface is developed to allow the individual to browse and extract knowledge from a large archive of lifelog data. Such archives could run for years or decades. This dashboard interface is built on top of a scalable framework that we hope to expand upon in future work. This first iteration of the design addresses specific lifelog datasets but it is highly adaptable. Different quantities and different varieties of data will require different and more innovative techniques for visualisation. For example, there is ongoing research into enhancing and improving the accuracy of the automatically detected lifelog concepts. Additionally, targeting specific behavioural change use cases via new dashboard features is also intended. These features could focus on aspects such as diet monitoring, frequency of exercise and exposure to other characteristics of healthy living. This interface and research will be demonstrated at the 12^{th} NTCIR conference in Tokyo this June as part of the Lifelog Insight Task (LIT) to explore knowledge mining and visualisation of lifelogs.

Acknowledgements

This paper is based on research conducted with financial support of Science Foundation Ireland under grant number SFI/12/RC/2289.

References

1. Dodge, M. and Kitchin, R. (2007). "Outlines of a world coming into existence": Pervasive computing and the ethics of forgetting. Environment and Planning, 34(3):431–445.

Laban Movement Analysis for real-time 3D gesture recognition

A. Truong, T. Zaharia

ARTEMIS Department, Institut Mines-Telecom, Telecom SudParis, CNRS UMR 8145 – MAP5, France.
{arthur.truong, titus.zaharia}@telecom-sudparis.eu

Abstract

In this paper, we propose a new method for body gesture recognition based upon Laban Movement Analysis (LMA). The features are computed for a dataset of pre-segmented sequences putting at stake 11 different actions, and are used to build a dictionary of key poses, obtained with the help of a k-means clustering approach. A soft assignment method based upon the obtained poses is applied to the dataset and assignment results are used as input sequences in a Hidden Markov Models (HMM) framework for real-time action recognition purpose. The high recognition rates obtained (more than 92% for certain gestures), demonstrate the pertinence of the proposed method.

Introduction

Gesture analysis is a highly challenging issue that involves both computer vision and machine learning methodologies. The emergence of general public, affordable depth cameras (e.g., Kinect) facilitating 3D body tracking can explain the recent growth of interest for gesture analysis. Gesture interpretation is useful for numerous applications: e-health, video games, artistic creation, video surveillance, immersive and affective communication. However, the issue of high level, semantic interpretation of gestures still remains a challenge, which requires the elaboration and development of effective gesture descriptors and recognition algorithms.

The approaches presented in (Huang & Guangyou, 2007) (Junejo, Junejo, & Al Aghbari, 2014) (Chu & Cohen, 2005) (Li, Zhang, & Liu, 2010) show that body poses are pertinent mid-level representation gestures structural aspects. Moreover, the possibility offered by the tracking of body joints provides new keys for mid-level features construction or pose extraction (Jiang, Zhong, Peng, & Qin, 2014) (Hussein, Torki, Gowayyed, & El-Saban, 2013) (Xia, Chen, & Aggarwal, 2012) (A Decision Forest Based Feature Selection Framework for Action Recognition from RGB-Depth Cameras, 2013). Still, the greatest part of the body motion features introduced are often dedicated to visual indices of motion. Thus, they usually fail to take into account the semantic aspects of motion (inter-subjectivity, expressivity, intentionality) and remain focused on gestures structural descriptions. The Laban Movement Analysis (LMA (Laban, 1994)), proposed by choreograph and dancer Rudolf Laban, provides a consistent representation of gestures expressivity. This framework has become a reference framework for different types of approaches. In our previous work (Truong, Boujut, & Zaharia, 2015), we have quantified several Laban qualities to directly exploit the features extracted for gesture recognition in a machine learning framework, without explicitly determining the underlying Laban components. In this paper, we introduce a new set of local descriptors based on LMA, aiming at characterizing each frame of a gestural sequence, and a mid-level representation of such features based on a soft assignment procedure that maps the LMA descriptors onto a reduced set of reference poses, automatically extracted from a learning gesture data. A HMM-based learning approach t exploits the soft assignment representation for real-time gesture recognition.

Proposed gesture descriptors

The *Laban Movement Analysis* (LMA (Laban, 1994)) model, introduced by Rudolf Laban consists of describing gestures in terms of *qualities* relating to expressivity and inter-subjectivity. These qualities relate to the intentional aspect of gestures, independently on the precise trajectory of the body in space. Laban system is composed of the following five major qualities: *Relationship, Body*, *Space*, *Effort*, *Shape*. The *Relationship* component refers to the relationships between individuals and is particularly suited in the case of group

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

performances. Thus, it is not required for designing of an intermediary gesture model which aims at characterizing various individual actions. The *Body* component deals with body parts usage, coordination and phrasing of the movement. The *Space* component refers to the place, direction and path of the movement. The *Effort* component depicts how the body concentrates its effort to perform the movement and deals with expressivity and style. The Effort further includes the following sub-components: *Space* (not to be confused with Space quality) that we did not retain in our work because of its redundancy with the generic *Space* component; *Time* which separates movements between sudden and sustained (or continuous) ones; *Flow* which describes movements as free or constrained; *Weight*, which aims at distinguishing between light and heavy movements. The *Shape* description is decomposed into three sub-components. *Shape flow* sub-component describes the dynamic evolution of the relationships between the different body parts. *Directional movement* sub-component describes the direction of the movement toward a particular point. *Shaping* sub-component refers to body forming and how the body changes its shape in a particular direction: rising/sinking, retreating/advancing and enclosing/spreading oppositions are respectively defined along the directions perpendicular to the horizontal, vertical and sagittal planes.

The proposed descriptors are based on 3D trajectories associated with the body skeleton joints that can be recorded with a depth sensor (*i.e.*, Kinect camera) at a rate of 30 frames per second. The Kinect sensor provides a maximum number of 20 joints, corresponding to the following body parts: *Center of the Hip*, *Spine*, *Center of the Shoulders*, *Head*, *Left Shoulder*, *Left Elbow*, *Left Wrist*, *Left Hand*, *Right Shoulder*, *Right Elbow*, *Right Wrist*, *Right Hand*, *Left Hip*, *Left Knee*, *Left Ankle*, *Left Foot*, *Right Hip*, *Right Knee*, *Right Ankle*, and *Right Foot*. Each body joint trajectory i is represented as a sequence of $(x_{i,t}, y_{i,t}, z_{i,t})_{t=0}^{N-1}$ coordinates in a 3D Cartesian system of coordinates (*Oxyz*) where *N* denotes the number of frames of the sequence. Several elementary transforms are applied to the body at each frame of its trajectory before the computation of descriptors. The objective is to set each body joint *i* in a new position at frame *t* $(x_{i,t}^{trans}, y_{i,t}^{trans}, z_{i,t}^{trans})$ so that the (*xOy*), (*yOz*) and (*zOx*) planes respectively correspond to sagittal, vertical and horizontal body planes. The result of this alignment process is shown on **Fig. 1**.

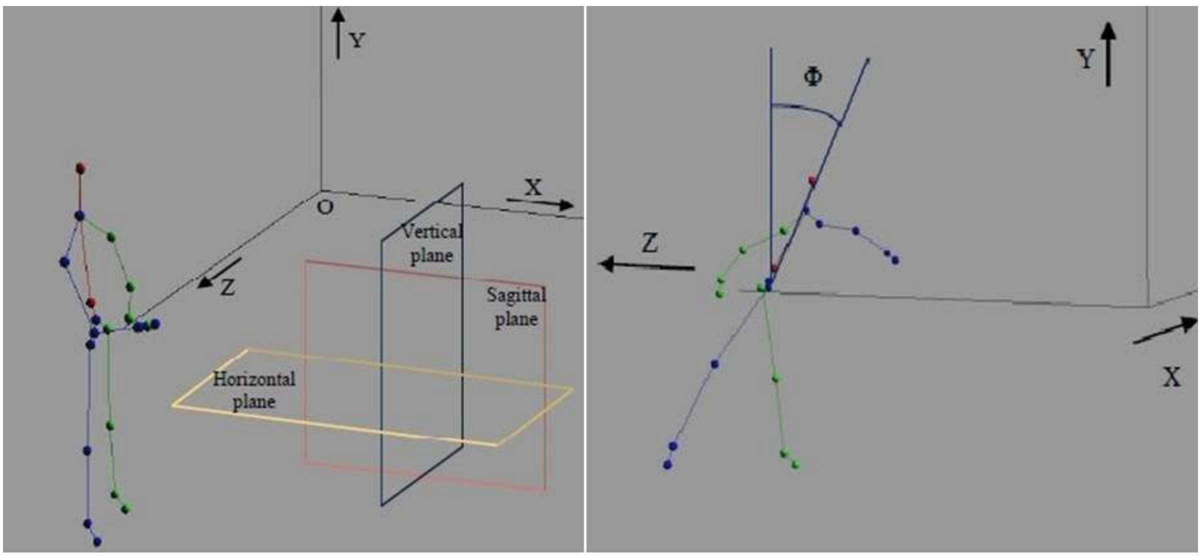


Fig. 1 Body skeleton joints new positions after transforms application and body planes representation (a); illustration of forward tilt angle (b).

Once these elementary transforms applied at frame t, a feature vector *v(t)* = $(v_1(t), \dots, v_P(t))$ can be computed for a total number of P = 17 features involving Space, Effort, Shape and Body quality.

The *Space* quality is described with the help of 2 values. The first one, defined as the *x* component of the head position $x_{Head,t}^{trans}$, characterizes the head forward-backward motion. The second value is the forward tilt angle

Φ(t) defined as the angle between the vertical direction *y* and the axis binding the center of the hip and the head, expressed in radians (**Fig. 1**).

The *Flow* sub-component of the *Effort* quality is described with the help of the third order derivative modules of the left and right hands trajectories, so-called jerk.

For the *Weight* sub-component of *Effort* quality, we consider the vertical components of the velocity and acceleration sequences (*i.e., $y'_{.,t}$* and *$y''_{.,t}$* signals) associated to 3 joints: the center of the hip, the left and the right hand. These 6 new values describe the vertical motion of the gesture.

The Shape flow sub-component of *Shape* is described by an index characterizing the contraction of the body, as defined in equation (1):

$$C(t) = \frac{(\left|x_{Left\,Hand,t}^{trans} - x_{Hip\,Center,t}^{trans}\right| - \left|x_{Right\,Hand,t}^{trans} - x_{Hip\,Center,t}^{trans}\right|)}{2}, \quad (1)$$

and has been inspired by the contraction index suggested in (Camurri, Lagerlöf, & Volpe, 2003).

Shaping sub-quality of *Shape* is quantified by 3 values corresponding to the amplitudes in the directions perpendicular to vertical, horizontal and sagittal planes (**Fig. 1**) respectively denoted by A_t^x, A_t^y, and A_t^z and defined by the following equations:

$$A_t^x = (max_i(\left|x_{i,t}^{trans}\right|) - min_i(\left|x_{i,t}^{trans}\right|)), \quad (2)$$

$$A_t^y = (max_i(\left|y_{i,t}^{trans}\right|) - min_i(\left|y_{i,t}^{trans}\right|)), \quad (3)$$

$$A_t^z = (max_i(\left|z_{i,t}^{trans}\right|) - min_i(\left|z_{i,t}^{trans}\right|)), \quad (4)$$

where *i* indexes the skeleton joints.

Finally, the *Body* component is quantified with the help of 3 features. The first one is an index characterizing the spatial dissymmetry between the two hands and has been inspired by the symmetry index proposed in (Glowinski, et al., 2011). This dissymmetry index is defined as described by the following equation:

$$Dys(t) = \frac{d_{left,\,center}(t)}{d_{left,\,center}(t) + d_{right,\,center}(t)}, \quad (5)$$

where $d_{left/right,\,center}$(t) denotes the distance between the left/right hand and the center of the shoulders. The Dys measure takes values within the [0, 1] interval. For a perfectly symmetric gesture, Dys equals 0.5. The second and third values are respectively the distance between left hand and left shoulder and the distance between right hand and right shoulder, respectively defined in equations (6) and (7). These parameters are used as a characterization of *Body* quality in (Aristidou & Chrysanthou, 2014):

$$D_t^g = \left|x_{Left\,Hand,t}^{trans} - x_{Left\,Shoulder,t}^{trans}\right|, \quad (6)$$

$$D_t^d = \left|x_{Right\,Hand,t}^{trans} - x_{Right\,Shoulder,t}^{trans}\right|, \quad (7)$$

Let us describe how such features can be exploited for actions recognition.

Gesture descriptors analysis

The proposed framework is illustrated in Fig. 2, with both off-line (learning) and on-line (classification) stages. We made the hypothesis that gestures characterization with the help of our descriptors would facilitate the capture of key poses, in order to build a dictionary. Thus for each gesture class G, we select the most representative poses.

Let us consider the gathering of all the instantiations of a gesture class G: $(S_1^G, S_2^G, \ldots, S_{|G|}^G)$, where |G| is their number. Each instantiation *i* corresponds to a series of frame descriptors: $S_i^G = (v^{G,i}(1), v^{G,i}(2), \ldots, v^{G,i}(T^{G,i}))$, where $T^{G,i}$ denotes the duration of the body motion *i*. We compute the key poses over all the sequences $(S_1^G, S_2^G, \ldots, S_{|G|}^G)$ by using classical K-medians clustering algorithm (Juan & Vidal, 2000) with random initialization (*i.e.*, the first 10 centroids are randomly selected among all the frames of all the sequences). We obtain a vector of poses for gesture G: $(P_1^G, P_2^G, \ldots, P_{10}^G)$. Each P_j^G kept in the dictionary consists of a skeleton pose and the feature vector corresponding to the descriptors computed at the chosen frame.

Once the class dictionaries built, all the poses are gathered into a unique dictionary. The size of the obtained dictionary is reduced by merging clusters whose centroids are too close. Thus, two clusters *C1* and *C2* whose centroids $\mu 1$ and $\mu 2$ are separated by a distance which is less than a certain threshold *th* will be merged. Finally, we obtain a vector of poses for the whole dataset: $(P_1^{GL}, P_2^{GL}, \ldots, P_M^{GL})$, where *M* is the size of the final dictionary set.

The references gathered in the global dictionary are relevant representatives of the trajectories run by the individuals in descriptors space and offer a key to simplify the representation of the gestures. The soft assignment method (Liu, Wang, & Liu, 2011) is commonly used to locate a feature vector among these type of representatives. For any feature vector $v(t) = (v_1(t), \ldots, v_P(t))$ at frame *t*, it is possible to compute the distance $d_j(t)$ of *v(t)* to every key pose P_j^{GL} of global dictionary: $d_j(t) = d\left(v(t), P_j^{GL}\right), j \in [1, M]$. Soft assignment vector *o(t)* will be composed of these distances normalized by their sum: $o(t) = (d_1'(t), d_2'(t), \ldots, d_M'(t))$ where $d_j'(t) = \frac{d_j(t)}{\sum_{i\in[1,M]} (d_i(t))}$, $j \in [1, M]$. *o(t)* gives the relative position of the vector in the space drawn by the key poses at frame *t*.

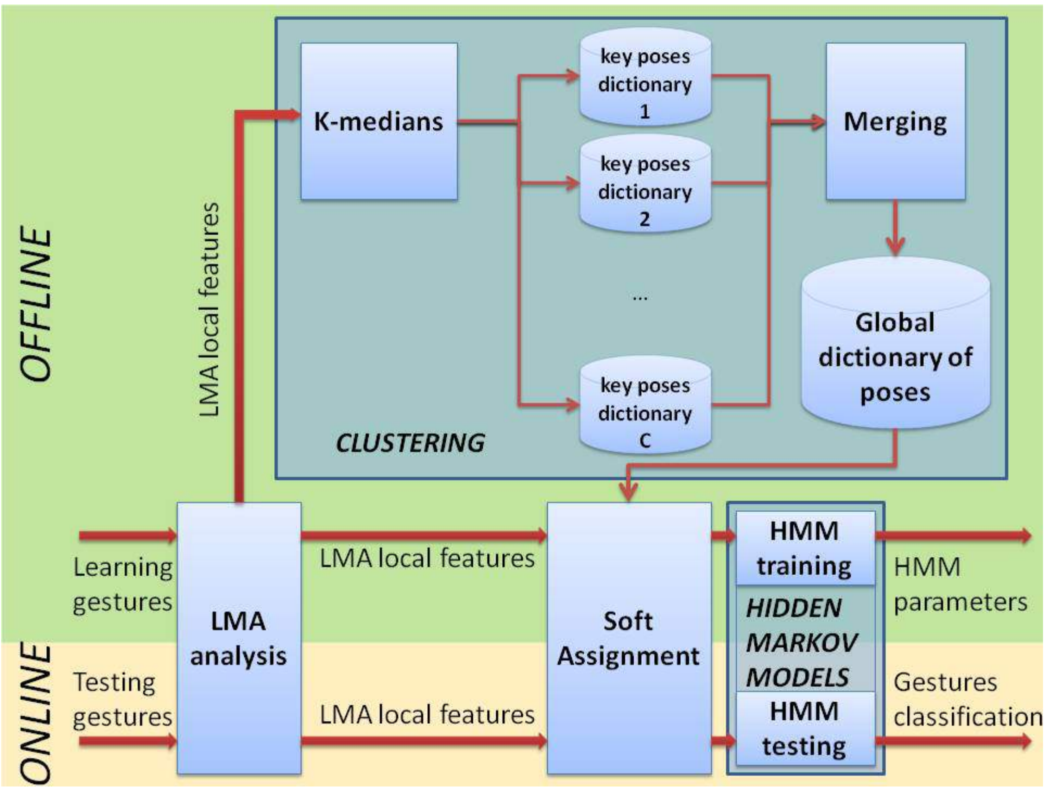


Fig. 2 Overview of the dynamic gesture analysis approach proposed.

Experiment and experimental results

In order to evaluate the ability of our descriptors model to dynamically characterize a gestural content, we designed a new dataset comprising 11 actions: *say "thank you" in American Sign Language*, *tie shoelaces*, *draw a circle with the right arm*, *rotate on oneself*, *catch an object*, *juggle*, *throw an object in front*, *cover one's ears*, *rub one's eyes*, *kneel*, and *stretch out*. We have finally obtained 565 gestures whose durations vary from 1s to 15s. Recording sessions were designed so as to preserve skeleton trajectories from occlusions.

We followed the protocol described in the previous section to our global dictionary. Then, we applied the soft assignment method to all gestures of all classes of the database. The soft assignment vector sequences obtained were used as observations in a classical Hidden Markov Models framework (HMM (Rabiner & Juang, 1986)) whose emission probabilities were modeled by Gaussian distributitons.

Let us denote by S the number of sequences and by $(\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_S)$ and $(\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_S)$ the observations sequences and the corresponding hidden states series.

We trained the HMM using Baum-Welch algorithm which consists of maximizing observations expectation: $(\sum_{s \in [1,S]} logP(\mathcal{O}_s|Model))$.

The decoding stage was performed with the help of Viterbi decoding procedure which maximizes a posteriori probability: $Best\ \mathcal{S} = arg\ max_{\mathcal{S}}(logP(\mathcal{S}|O, Model))$.

At the testing stage, every observation sequence was decoded as a succession of gesture labels taking their values in the list of our 11 actions. For each decoding sequence, we decided to keep the 3 classes which were the most decoded. Then, for every gesture class C, we computed the recognition rates in first, second and third positions to evaluate the consistency of our system.

The obtained results are summarized in Fig. 3.

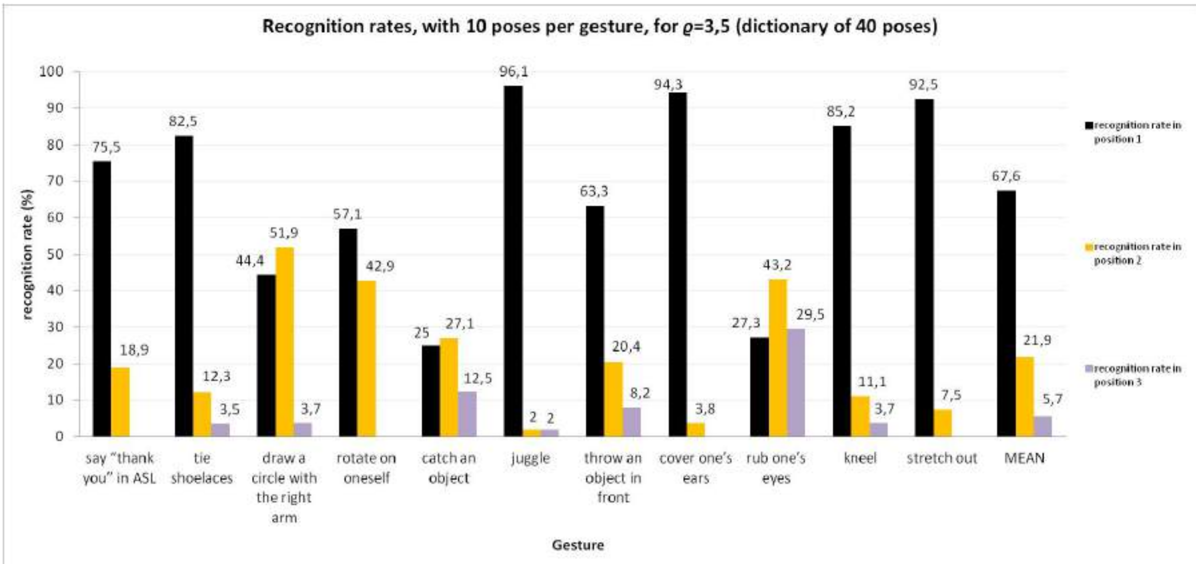


Fig. 3 Recognition rates per class for HTI 2014-2015 dataset.

The best recognition results are obtained for a dictionary of 40 poses. We can see that *tie shoelaces*, *juggle*, *cover one's ears*, *kneel*, *stretch out* gestures yield recognition rates greater than 80%. 3 gestures yield recognition rates over 90%: *juggle*, *cover one's ears* and *stretch out*. *say thank you* and *throw an object in front* gestures give mixed results.

We also tested our method on other available datasets. Fig. 4 and Fig. 5 present the results respectively obtained on iconic gestures subset and metaphoric gestures subset of Microsoft Gesture dataset (MSRC-12 dataset, which we had used in our previous work (Truong, Boujut, & Zaharia, 2015)). Our approach yields recognition rates close to the ones obtained on the same corpus (Song, Morency, & Davis, 2013) (A Decision Forest Based Feature Selection Framework for Action Recognition from RGB-Depth Cameras, 2013) (Hussein, Torki, Gowayyed, & El-Saban, 2013) (Zhao, Li, Pang, & Zhu, 2013), excerpt for certain metaphoric gestures: *start system*, *wind it up* and *beat both* (the worst score is reached for this last category, with 39.1% recognition rate). We also tested our method on UTKinect-Human Detection dataset which is exploited in (Xia, Chen, & Aggarwal, 2012). The results in Fig. 6 shows that half of the gestures reach recognition rates greater than 95%. There is only one category whose score is less than 70%, which shows the relevance of our approach.

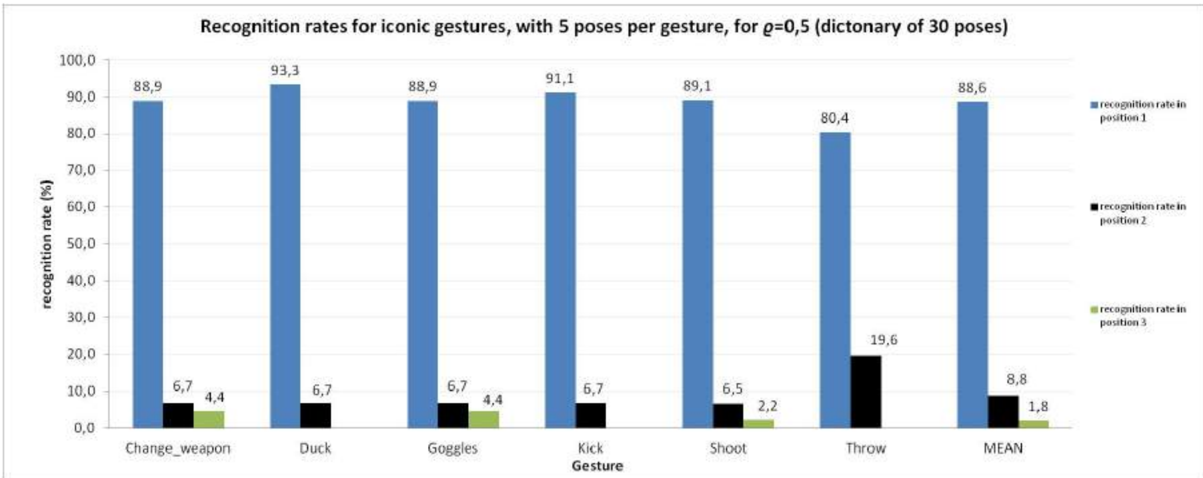


Fig. 4 Recognition rates for iconic gestures of MSRC-12 dataset.

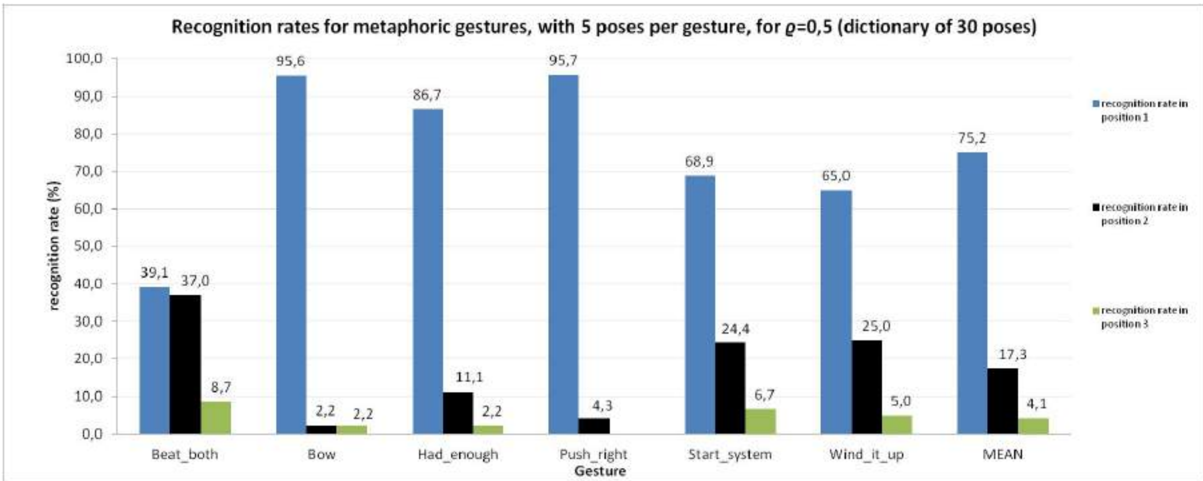


Fig. 5 Recognition rates for metaphoric gestures of MSRC-12 dataset.

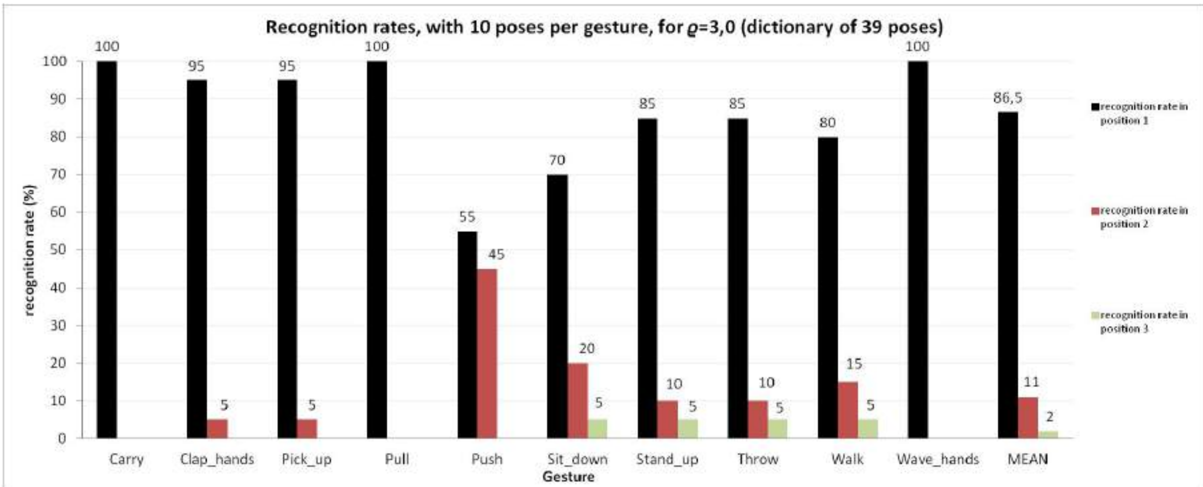


Fig. 6 Recognition rates per gesture for UTKinect Human Detection dataset.

Conclusion and perspectives

In this paper, we introduced a gesture analysis approach, based on a set of local descriptors dedicated to the various entities defined in the LMA framework. Our approach yielded high recognition rates on different datasets of actions. Our perspectives of future work concern the testing of our approach on sequences of actions, or in other frameworks, like gaits, affective states or emotions analysis.

References

1. Feiyue Huang and Xu Guangyou, "Viewpoint insensitive action recognition using envelop shape," in *Computer Vision-ACCV 2007*.: Springer Berlin Heidelberg, 2007, pp. 477-486.
2. Imran N. Junejo, Khurrum Nazir Junejo, and Zaher Al Aghbari, "Silhouette-based human action recognition using SAX-Shapes," The Visual Computer, vol. 30, no. 3, pp. 259-269, 2014.
3. Chi-Wei Chu and Isaac Cohen, "Posture and gesture recognition using 3D body shapes decomposition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops*.: IEEE, 2005, pp. 69-78.
4. Wanqing Li, Zhengyou Zhang, and Zicheng Liu, "Action recognition based on a bag of 3d points," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.: IEEE, 2010, pp. 9-14.
5. Xinbo Jiang, Fan Zhong, Qunsheng Peng, and Xueying Qin, "Online robust action recognition based on a hierarchical model," *The Visual Computer*, vol. 30, pp. 1021-1033, 2014.
6. Mohamed E. Hussein, Marwan Torki, Mohammad A. Gowayyed, and Motaz El-Saban, "Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations," in *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*.: Press, AAAI, 2013, pp. 2466-2472.
7. Lu Xia, Chia-Chih Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3d joints," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.: IEEE, 2012, pp. 20-27.
8. Farhood Negin, Firat Özdemir, Ceyhun Burak Akgül, Kamer Ali Yüksel, and Aytül Erçil, "A Decision Forest Based Feature Selection Framework for Action Recognition from RGB-Depth Cameras," in *Image Analysis and Recognition*.: Springer, 2013, pp. 648-657.

9. Rudolf Laban, La Maîtrise du Mouvement. Arles: Actes Sud, 1994.
10. Arthur Truong, Hugo Boujut, and Titus Zaharia, "Laban descriptors for gesture recognition and emotional analysis," *The Visual Computer*, pp. 1-16, 2015.
11. Antonio Camurri, Ingrid Lagerlöf, and Gualtiero Volpe, "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 213-225, 2003.
12. Donald Glowinski et al., "Toward a minimal representation of affective gestures," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 106-118, 2011.
13. Andreas Aristidou and Yiorgos Chrysanthou, "Feature Extraction for Human Motion Indexing of Acted Dance Performances," *GRAPP 2014 - International Conference on Computer Graphics Theory and Applications*, 2014.
14. Alfons Juan and Enrique Vidal, "Comparison of four initialization techniques for the k-medians clustering algorithm," in *Advances in Pattern Recognition*.: Springer, 2000, pp. 842-852.
15. Lingqiao Liu, Lei Wang, and Xinwang Liu, "In defense of soft-assignment coding," in *2011 IEEE International Conference on Computer Vision (ICCV)*.: IEEE, 2011, pp. 2486-2493.
16. Lawrence R. Rabiner and Biing-Hwang Juang, "An introduction to hidden Markov models," *ASSP Magazine, IEEE*, vol. 3, no. 1, pp. 4-16, 1986.
17. Yale Song, Louis-Philippe Morency, and Randall Davis, "Distribution-Sensitive Learning for Imbalanced Datasets," in *2013 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2013)*.: IEEE, 2013.
18. Xin Zhao, Xue Li, Chaoyi Pang, and Xi Zhu, "Online human gesture recognition from motion data streams," in *Proceedings of the 21st ACM international conference on Multimedia*.: ACM, 2013, pp. 23-32.

Emotional Assessment and Stress Management Through Biofeedback Training: Are Wearable Sensors and Devices Ready for Making People Learning How to Feel Better?

Maurizio Mauri[1], Vincenzo Russo[1], Otello Sorato[2]

[1]IULM University of Milan, Italy
[2]Mind Room Lab, Italy

Introduction

In this work it is presented an experimental study carried out at the "Behavior & Brain Lab" of the IULM University of Milan. The experiment was realized applying both traditional and neuroscientific methods to evaluate the effects of a training about stress management. Neal Miller in the fifties introduced the biofeedback techniques to empower the ability to affect mental states (Miller, 1978). Further studies, although there was still a resistance in accepting the role of cognition in mental studies, showed that the learning of specific motor abilities is bounded to the ability to create mental models (called neuro-motor programs) about how the specific motor behavior should be accomplished. Especially when it could be possible to learn by means of a mere observation and then imitation of the model (Rosenthal and Zimmerman, 1978). The possibility for people to see a continuous indicator about the performance can enhance the level itself of the performance. This is the scientific rationale about the biofeedback training. In particular, the possibility to voluntarily control and regulate the respiration can change and reduce other physiological indicators (Jenis, De Fare, Grossman, 1983; Tymmons and Ley, 1994). Although there are many studies supporting the efficiency of biofeedback techniques, it is still hard to teach people how to take advantage of this scientific knowledge. In this work an experimental study about the effects of biofeedback is presented, and implications for application in everyday life by means of wearable technologies is discussed.

Experimental procedures

Twenty students (total n = 20) voluntarily joined the study after filling in a general self-report about their health (no cardiac or psychological disease and/or treatment at the time of the study) and purpose of the study. Two groups, both composed by 10 PhD students from different fields (physics, medicine, psychology, engineering, Italian literature, etc.), were asked to undergo different measurements: psychological self-reports, and psycho-physiological assessment. The psychological assessment was represented by 3 different self-reports: the State Trait Anxiety Inventory (STAI), the Anxiety for Social Relation Inventory (SIAS), mainly related to the ability to manage stress and anger, and the Emotional Regulation Questionnaire (ERQ). The physiological assessment consisted in the monitoring of several indexes (such as heart rate, skin conductance levels, respiration, electromyography of shoulder muscles and peripheral temperature, all recorded by a FlexComp device from Thought Technology) during different mental activity, such as: relax, mental math calculation, relax, stroop-test, reading, relax. All study participants underwent the psychological and physiological assessments twice: the first time at the beginning of the research (T0), and the second one at the end of the study (T1). Between T0 and T1 (2 weeks), the participants from one group, called the "experimental group", were exposed to 5 sessions of biofeedback training. The other group, who was not exposed to the biofeedback training, has been called "control group". An ethical committee composed by professors in the field of General Psychology and Work Psychology form IULM University approved the experiments.

Results

Results from T0 assessments and T1 reassessment were compared in order to evaluate the effects of the biofeedback training. The scores from the self-reports and the means values for all indexes have been calculated. For the State Trait Anxiety Inventory, there were significant differences between T0 and T1 for the experimental

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

group, who has been exposed to the training course, but not for the control group, who has been not exposed to the biofeedback training group. Also about physiological indexes for respiration and skin conductance there were significant differences between T0 and T1 for the experimental group only. No significant difference were found in the control group, in terms both of psychological self-reports and psychophysiological indicators.

Conclusion

Results shows how the biofeedback training enabled participants in empowering their ability to reduce their level of anxiety after a stress exposure (stroop task), both in terms of psychological self-reports and psycho-physiological indexes. As today there are many devices as bracelets and belts that can monitor cardiorespiratory indexes and skin conductance, it should be possible to allow people to take advantage of these findings. However, the role of technology is essential in enabling people to understand their emotional reactions associated to their physiology. Implications, limits and further directions in this fields will be discussed. The ideal way to apply this findings into real life outside the labs' walls would be to create wearable sensors and portable devices (smartphones) with application (app) that could help anybody in taking advantage of the knowledge about psychophysiology today available. The app could guide people in increasing their emotional intelligence (as the "brain training" does) and their wellness.

References

1. Miller, N.E. (1978), "Biofeedback and Visceral learning", *Annual Review Psychology*, 29: 373-404.
2. Rosenthal, T.L. and Zimmerman, B.J. (1978), *Social Learning and Cognition*, Academic Press Inc.
3. Janis, I.L., Defares, P.B. and Grossman, P. (1982). *Hyper vigilant reactions to threat. In H. Selye H.* (Ed.) Selye's guide to stress research, 3, Scientific and Academic Editions, New York.
4. Tymmons, B.H. and Ley, R. (1994). Behavioral and Psychological Approaches to Breathing Disorders, Plenum Press, New York.

Measuring Activities of Daily Living: Digitally, Visually and Semantically

Peng Wang

Department of Computer Science and Technology, Tsinghua University, Beijing, China.
pwang@tsinghua.edu.cn

Abstract

Accurate measurement of activity of daily living (ADL) has many benefits in different aspects such as in analyzing human lifestyle, diet monitoring, occupational therapy, active rehabilitation, etc. Traditionally self-report approaches to measuring ADL have been developed to assess the engagement in meaningful activities. Given the inaccuracy nature of self-report approaches limited by the knowledge and objectivity of the respondents, observational assessment have been widely used to assess ADL to further establish care needs according to observed performances. Because frequent in-situ observations for ADL measuring are time consuming which are limited by time constraints of the assessor, the measuring of ADL requires efficient solutions to help to localize meaningful activities from large volume digital archives collected continuously with sensory devices. The wildly adoption of mobile or wearable devices makes it feasible to measure everyday activities digitally, with their built-in sensing and computational capabilities. This paper investigates the state-of-the-art developments in semantic learning from visual media captured by wearable devices, and explores how these techniques can be incorporated in the proposed solution to support accurate ADL measurement and improve activity engagement.

Keywords: activity of daily living, wearable sensing, visual media, semantic learning, visual lifelogging

Introduction

While more concerns are shown in modern life on human health, growing body of evidence has been demonstrated indicating the relationship between time spend and well-being within various age groups[1,2]. Some studies also show that engagement in everyday occupations (instrumental, leisure, social) is associated with health and life satisfaction [1,2]. In [3,4], studies have shown relationships between important health outcomes and sedentary behavior, independent of physical activity.

Traditionally self-report methods have been used to understand an individual's activity engagement and routines. Standard tools have been developed to support client-centered approaches to assessing and improving engagement in meaningful activities. These tools include interview-based or questionnaire-based measures to reflect frequency of participation. It is known that there are errors according to the results reported by these tools employed in behavior measurement. Though self-report methods like questionnaires or diary interview can help identify specific behavior types these methods are constrained and plagued by memory recall errors, such as omission and telescoping, and other forms of bias associated with comprehension and positive representation [3,5].

Since accurate measurement of the activity engagement is important, researchers turn to the employment of digital devices to automatically record sensor data in order to reflect contexts of physical activity episodes in more detailed granularity. Thanks to the characteristics of portable or even wearable devices like light-weight and long-battery life, the application of these digital sensing devices opens new directions in automatically measuring human behavior. For example, physical activities are detected from data acquired using five accelerometers worn simultaneously on different parts of the body in [6]. Similarly in [7], activity recognition is explored from tri-axial accelerometers to model a user's activity as a means of enabling context-aware retrieval. According to [8], accelerometers can identify certain physical activity behaviors, but not the context in which they take place. The feasibility of wearable cameras to objectively categorize the behavior type and context of participants' accelerometer-identified episodes of activity is then investigated by [8]. The results have shown that wearable cameras represent the best objective method currently available to categorize the social and environmental context of accelerometer-defined episodes of activity in free-living conditions. In [9], a study is also carried out to test the feasibility of using wearable cameras to generate, through image prompted interview, reconstructed 'near-objective' data to assess the validity of the records of daily activities. In their study the feasibility of using visual wearable cameras to reconstruct time use through image prompted interview is demonstrated.

Though visually measuring activities is shown to be promising, there still lacks of automatic means to identify meaningful activities from large volume of visual captures. In [10], a technique for automatically identifying the activity type of events is proposed using Support Vector Machine (SVM) learning of Speeded-Up Robust Features (SURF) visual words (a codebook of 4,000 words) extracted from all the SenseCam images. Though 23 concepts can be automatically detected for semantic indexing of visual captures, the identification of everyday activities is limited without modeling the time-varying patterns of activity evolution. The use of state-of-the-art semantic indexing technologies to infer high-level activity understanding is still necessary for flexible behavior measuring. Meanwhile, the automatic detection of semantic concepts from images has now reached a usable performance level, the utilization of everyday concepts captured by the imagery to temporally characterize everyday activities is proposed and discussed in the proposed solution by the following sections.

Devices and Proposed Solution

The SenseCam, shown in Fig. 1, is a sensor-augmented wearable camera designed to capture a digital record of the wearer's day by recording a series of images and a log of sensor data. It captures the view of the wearer from a fisheye lens and pictures are taken at the rate of about one every 50 seconds without the trigger of other sensors. The on-board sensors for measuring ambient light levels, movement, and the presence of other people through a passive infra-red sensor, are also used to trigger additional capture of pictures when sudden changes are detected in the environment of the wearer as well as to prevent images being captured when the wearer, and the SenseCam, are being moved which would result in blurring of images. Due to its advantages of multiple sensing capabilities, light weight and unobtrusive logging with a long battery life, we employ SenseCam as a wearable device to discuss details of the proposed solution. The method discussed in this paper is also applicable to the employment of other wearable cameras like Narrative, Autographer, etc. Such continuous recording of every aspects of everyday activities using visual cameras can be termed as visual lifelogging.

Fig. 1 The Microsoft SenseCam (right as worn by a user).

The state-of-the-art approaches to identifying semantics from visual media use statistical techniques to map low-level local or global features like color, texture or shape, to high-level semantic concepts like "indoor", "building" or "walk", a process termed "concept detection". The natural progression is from a lifelog image or video, to a set of such semantic concepts occurring in the image, and then to infer an activity the wearer was participating in while the image was taken based on the presence or absence of semantic concepts. Following that, we can then to aggregate activities over a long period of time in order to reason about the wearer's behavior or identify changes in it.

However, despite recent progress, automatic concept detectors are still far from perfect and how to classify high-level activities based on such noisy semantic attributes needs to be tackled. This is especially important for cases where we then build upon the detected concepts such as using them to infer activities and then behavior. For measuring behavior applications, there is a further challenge because of the diverse range of usable concepts, and the generally noisy nature of the lifelog data because of the wearers' movements and because even the images captured passively within the same lifelogged activity may have significant perceptual differences.

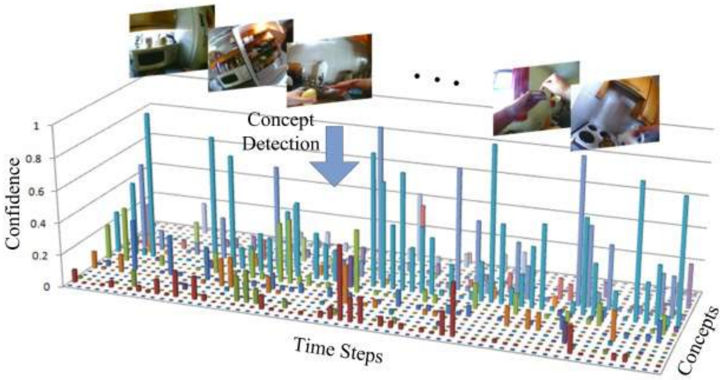


Fig. 2: The dynamics of concept attributes quantified by confidences returned by concept detections.

Inspired by recent work on concept-based temporal modeling [11,12], we model the dynamic evolution of human activities using concept detection results as input. In effect this means that streams of activities are represented as sequences of units such as clips or frames. Concept detections are applied to each of units and by concatenating the output results (confidences) of pre-trained concept detectors at the same time step as a vector, one activity stream can be represented by a temporally ordered sequence of vectors, as shown in Fig. 2.

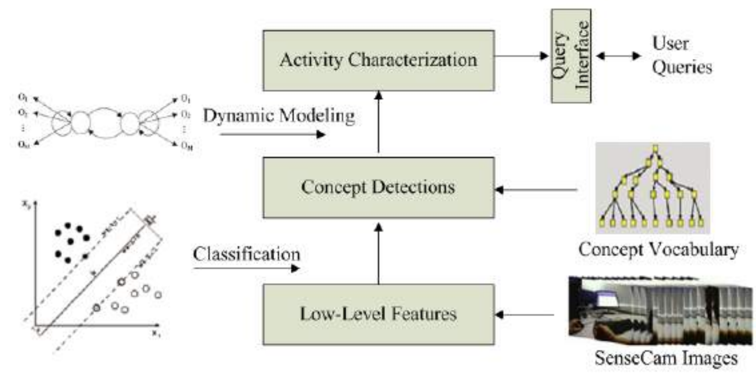


Fig. 3: Solution pipeline for everyday activity measurement.

The methodology we proposed for the semantic measuring of everyday activities can be demonstrated by the algorithm pipeline shown in Fig. 3. The solution consists of three main components which are low-level feature extraction from raw SenseCam images, concept detection based on semantic vocabulary, the modeling of time-varying dynamics to characterize activities. The bridging of low-level features and semantics with the fusion of concepts is to provide better understanding of everyday activities, which is known as a concept-based modeling approach to behavior measuring. In this approach, concepts are first detected by a mapping from low-level features using generic classification methods. The concepts are then fused temporally to model the dynamics of visual semantics in activity samples, which can then be queried through interfaces for more complex measuring applications.

In the solution as demonstrated in Fig. 3, the problem of activity characterization from lifelog images taken by a SenseCam wearable camera, can also be simplified into a classification problem, that is, we find the most likely activity type for an sample stream from a lexicon set with regard to the input. This classification task is indeed an automatic annotation of the presence or absence to a temporal period of SenseCam captures. The typical methods in concept-based activity characterization can be categorized into two classes: discriminative and generative methods. In [13] and [14], discriminative SVMs are applied on the fused features extracted based on Max-Pooling, which chooses the maximum confidence from all keyframes for each concept to generate a fixed-dimensional vector for an activity. Similar discriminative methods are applied in [15,16,17] using Bag-of-Features which aggregates concept detection results by averaging the confidences over time window. Fisher Vector is another typical discriminative features which describes how the parameters of the activity model should be modified in order to adapt to different samples [11,18]. Because the presence or absence of concept attributes is likely to show patterns in time, temporal models such as generative HMMs [12] have the advantage of capturing the temporal dependency of nearby states. Since the purpose of this paper is to propose a solution based on concept attributes rather than focusing on specific modeling of such attributes, more details of temporal modeling of visual semantics for everyday activities can be found in [12, 18, 19].

Challenges in Semantic Activity Measuring

Concept Detection Accuracies

In [12], the correlation of activity and concept detection performances is analyzed and results have shown that improving concept detection accuracies, the recognition accuracies of everyday activities can be enhanced accordingly. Given the fact that current state-of-the-art concept detectors are still far from perfect despite recent progress, improving original concept detections for enhanced activity recognition is one challenge in semantic activity measuring. Another potential solution is to combine concepts or post-processing the outputs of individual concept detections returned by independent concept detectors. This is motivated by the intuition that inter-concept relationships or dependencies exist in many concept pairs and triples which usually co-occur rather than occur independently. Current research in concept detection refinement has demonstrated the effectiveness of this intuition in lifelogging domain and detection accuracy for concepts can be improved if concept correlation can be exploited [19,20,21,22].

More specifically, the quality of visual media captured by wearable devices is another affecting factor to concept detection accuracies. As shown in Fig. 3, low-level features should be extracted from raw SenseCam images in the proposed solution. However, the poor quality of raw visual media captured by wearable devices is an intrinsic challenge. Taking SenseCam as an instance, though the on-board sensors are employed to alleviate the problem of image blurring, the images are still suffering from motion blur when captured while the wearer are being moved, as shown by the first row in Fig. 4. In addition, since SenseCam is worn around the neck by the user while collecting the data, the lens is often blocked by clothes or even arms. This will cause images with a narrower visual field as shown by the first two samples in the second row of Fig. 4. Overexposure (the last image, second row in Fig. 4) is another quality issue making it difficult to automatically categorize and index the media.

Fig. 4: Typical image quality issues faced in visual activity measurement.

To tackle the above image quality issues, a pre-processing step can be first applied before low-level feature extraction in which low-quality images should be filtered out first according to the quality metrics. For example, a fusion of the Contrast and Saliency measures can be applied and this has shown to be effective in choosing high-quality everyday activity representations [23,24]. Note that the tradeoff of data volume and image quality also leads to low image capture rate. This makes dynamic descriptors, spatial-temporal features like HOG (Histograms of Oriented Gradients) and HOF (Histograms of Optical Flow) descriptors, inapplicable, which are well adopted in traditional video classification.

Concept Selection

The semantic method for everyday activity measuring is based on indexing wearable sensory media with a predefined lexicon as shown in Fig. 3 (denoted as concept vocabulary). This technique enables semantic reasoning and dynamic modeling on the concept set to facilitate high-level activity characterization and measures. To bridge the sensory gap between raw visual data and user expectations, a set of concept detectors is usually developed to represent the high-level metadata. In order to provide satisfactory activity recognition performance, the problem of mapping ambiguity between everyday activity and concepts needs to be solved.

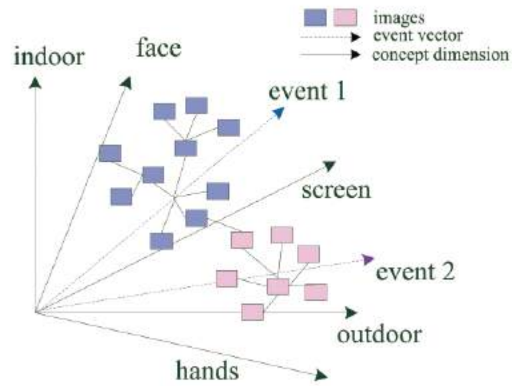


Fig. 5: Concept space and event/activity concept vector.

To select concepts to represent the semantics for events/activities, we need to define the concept space. Intuitively, every concept representing any activity should be one dimension, and the projection of an activity onto the concept space is the co-occurrence information in between. However, different concepts have different impacts on activity interpretation. Concepts which are neither too general nor too specific should be selected in the semantic space to reduce dimensionality and noise for concept detection. In a nutshell, we should include topic-related concepts with decent frequency, and exclude general and over-specific senses. The Event Semantic Space (ESS) is defined as a linear space with a set of concepts as the basis as depicted in Fig. 5. In order to ensure high coverage of the space, the selection of a minimum concept basis set is necessary according to the generalization of entities in the semantic space.

The benefits of appropriate concept selection are demonstrated in experiments as reported in [25]. The sophisticated selections of more appropriate concepts have shown to be effective for a better performance in activity recognition. For the lifelog activity indexing, topic-related semantic concept selections are investigated in [26] to choose concepts in terms of user experiments and semantic networks.

Dynamic Modeling

Besides the above discussed factors, dynamic modeling methods also play the dominant role in obtaining different performances in activity recognition. Therefore, how to better classify high-level activities streams based on noisy concept detection results needs to be tackled. This is especially important for cases where we then build upon the detected concepts such as using them to infer activities, complex events or behaviors. For such applications, there may be further challenges because of the diverse range of usable concepts, or the noisy nature of the visual data.

As described in Section 2, there are two main categories of activity characterization methods, which are discriminative and generative respectively. The state-of-the-art approach is to apply discriminative machine learning algorithms such as SVMs to decide the most likely activity class given the extracted features [14]. Compared to a discriminative model which is more task-oriented, generative statistical models such as Markov model try to analyze the joint probability of variables, which are also proposed in activity characterization [12]. Both generative and discriminative approaches have their own pros and cons. A generative model is a full probabilistic model of all variables whereas a discriminative model has limited modeling capability. This is because a discriminative model provides a model only for the target variable(s) conditional on the observed variables hence cannot generally express more complex relationships between the observed and target variables. However, discriminative models are often easier to learn and perform faster than generative models. Besides, it has been shown that discriminative classifiers often get better classification performance than generative classifiers with large training volume (usually including positive and negative samples), such as shown in [19] and [27].

Efficiency and Real-Time Concerns

While recognition accuracy is the main concern in current research on everyday activity measurement, we point out that the recognition efficiency should also be taken into account when designing real-world applications of measuring behavior. The activity analysis algorithms proposed for various purposes like life pattern learning, diet monitoring, memory aid, etc. are indeed carried out in a post-processing manner. That is, collecting digital records of everyday activities first and then applying analyzing methods on them offline. However, this limitation constrains further adoption of such technical schemes to real-time applications like context-aware personalized recommendation, smart home, tourism guide, etc.

To tackle the challenge of efficiently managing the high volume and diversity of activity data, distributed data processing strategy is one possible solution. In the distributed architecture like [28,29], the elastic nature of cloud infrastructure can be utilized to keep a tradeoff between performance and cost. For example, the mobile-side device such as wearable sensing is used to gather activity context information from heterogeneous sensors. The server side is responsible for data storage and apply high computational complexity software to enable the semantic recognition of everyday concepts and activities. Based on such architecture, more complex web application can be built efficiently though interfaces which allows semantic queries and result displays, such as in a semantic enriched visualization [30].

Not only the distributed architecture can be useful in developing measuring behavior as a real-time or even pro-active solution to encourage adaptive recommendation, healthcare warning, etc., the fusion of heterogeneous senor readings can also tackle the scalability and performance challenges of real-time analysis. In behavior measuring, most wearable devices such as SenseCam are power-strapped. The limited computational capability on mobile and wearable devices leads to memory constraints in recognizing activities quickly or transmitting large amount of sensor data to cloud side. An alternative strategy is useful to substitute the high power/memory consuming sensor by a lower one when the lower power/memory consuming sensor data can be fused and maintain comparable activity recognition accuracy. Such sensor data can be chosen from a large variety of contexts including environment information (light, temperature, pressure, etc.), bio-information (heart rate, galvanic response, etc.) and spacial information (location, acceleration, co-presence, etc.), according to different application requirements. These contexts are changing dynamically with activity evolution and if fused with visual semantics then they can be used as cues to our activities and thus help with accessing information in large volume behavior archives.

Conclusions

While measurement of activity of daily living can facilitate a variety of real-world applications such as analyzing human lifestyle, healthcare, personalized recommendation, etc., traditionally self-report and observational approaches to measuring activities suffer from limitations of accuracy, knowledge and objectivity of the respondents, as well as time expense of the assessor. The wildly adoption of mobile or wearable devices makes it feasible to measure everyday activities digitally, with the built-in sensing and computational capabilities. This paper investigates the state-of-the-art developments in semantic learning from visual media captured by wearable devices, and explores how these techniques can be incorporated in the proposed solution to support accurate ADL measurement and improve activity engagement. The challenges in designing practical semantics-based activity measuring applications using wearable sensing are also stressed and discussed, which are potential future research directions.

Acknowledgement

This work was supported by National Natural Science Foundation of China under Grant No. 61272231, 61472204, 61502264 and Beijing Key Laboratory of Networked Multimedia.

References

1. Law, M., Steinwender, S., Leclair, L. (1998). Occupation, health and well-being, *Canadian Journal of Occupational Therapy* 65 (1998) 81–91.
2. McKenna, K., Broome, K., Liddle, J. (2007). What older people do: time use and exploring the link between role participation and life satisfaction in people aged 65 years and over, *Australian Occupational Therapy Journal* 54 (2007) 273–284.
3. K., Jacqueline, Marshall S.J. et al (2013), Using the SenseCam to Improve Classifications of Sedentary Behavior in Free-Living Settings, *American Journal of Preventive Medicine*, 44(3), 2013, 290-296
4. Thorp, A.A., Owen, N., Neuhaus, M., Dunstan, D.W. (2011). Sedentary behaviors and subsequent health outcomes in adults: a systematic review of longitudinal studies, 1996–2011. *American Journal of Preventive Medicine*, 2011, 41(2):207–15.
5. Sallis, J. F., Saelens, B.E. (2000). Assessment of physical activity by self-report: status, limitations, and future directions. *Research quarterly for exercise and sport* 2000;71(2):1–14
6. Bao, L., Intille, S.S. (2004). Activity Recognition from User-Annotated Acceleration Data Ferscha, A., Mattern, F. (Eds.) *Pervasive Computing* 2004, 1–17.
7. Qiu, Z. W., Doherty, A.R., Gurrin, C., Smeaton, A.F. (2011). Mining user activity as a context source for search and retrieval, *International Conference on Semantic Technology and Information Retrieval (STAIR)*, pp.162-166, 28-29 June 2011
8. Doherty, AR, Kelly, P, Kerr, J, et al (2013). Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. *The International Journal of Behavioral Nutrition and Physical Activity*. 2013, 10:22.
9. Kelly, P, Thomas, E, Doherty, A, Harms, T, Burke, Ó, Gershuny, J, et al. (2015). Developing a Method to Test the Validity of 24 Hour Time Use Diaries Using Wearable Cameras: A Feasibility Pilot. *PLoS ONE* 10(12): e0142198.
10. Doherty, A. R., Caprania, N., Conairea, C.O., Kalnikaited, V. et al (2011). Passively Recognising Human Activities through Lifelogging, *Computers in Human Behavior*, 27(5), 2011, pp. 1948–1958
11. Sun C., Nevatia, R.. (2013). ACTIVE: Activity Concept Transitions in Video Event Classification. *International Conference on Computer Vision*, 2013. 913–920.
12. Wang, P., Smeaton, A.F.. (2013). Using visual lifelogs to automatically characterize everyday activities. *Information Sciences* 230, (2013), 147–161
13. Liu, J., Yu, Q., Javed, O., Ali , S., Tamrakar , A., Divakaran , A., Cheng , H., Sawhney, H. (2013). Video event recognition using concept attributes. *Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 339-346.
14. Bhattacharya, S., Kalayeh, M., Sukthankar, R., Shah, M.. Recognition of complex events exploiting temporal dynamics between underlying concepts. *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014.
15. Pirsiavash, H., Ramanan, D. Detecting activities of daily living in first-person camera views. *IEEE*

Conference on Computer Vision and Pattern Recognition, pages 2847-2854 , June 2012.

16. Merler, M., Huang, B. , Xie, L., Hua, G., Natsev, A. (2012). Semantic model vectors for complex video event recognition. *IEEE Transactions on Multimedia*, 14(1):88-101, Feb 2012.
17. Guo, J., Scott, D., Hopfgartner, F., Gurrin, C. (2012). Detecting complex events in user-generated video using concept classifiers. *10th Workshop on Content-Based Multimedia Indexing*, pages 1-6, June 2012.
18. Wang, P., Yang, S.Q., Sun, L.F., Smeaton, A.F. (2015). Improving the classification of quantified self activities and behaviour using a Fisher kernel. *In: UbiComp/ISWC'15 Adjunct*, 7-11 Sept 2015.
19. Wang, P., Yang, S.Q., Sun, L.F., Smeaton, A.F. (2015). Characterizing Everyday Activities from Visual Lifelogs based on Enhancing Concept Representation. *Computer Vision and Image Understanding*, 2015, DOI: 10.1016/j.cviu.2015.09.014
20. Wang, P., Smeaton, A.F., Zhang, Y.C., Deng, B. (2014). Enhancing the detection of concepts for visual lifelogs using contexts instead of ontologies, *International Workshop on the Visualization of Heterogeneous Multimedia Content, In conjunction ICME* 2014, Chengdu, China
21. Wang, P., Smeaton, A.F., Gurrin, C. (2015). Factorizing time-aware multi-way tensors for enhancing semantic wearable sensing, *21st International Conference on Multimedia Modeling*, 2015, pp. 571-582.
22. Wang, P., Yang, S.Q., Sun, L.F., Smeaton, A.F. (2015). Towards Training-Free Refinement for Semantic Indexing of Visual Media, *International Conference on Multimedia Modeling 2016*, Miami, Fl, USA.
23. Doherty, A.R., Byrne, D., Smeaton, A.F., Jones, G.J., Hughes, M. (2008), Investigating keyframe selection methods in the novel domain of passively captured visual lifelogs, *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval, CIVR'08*, 2008, pp. 259-268.
24. Wang, P., Smeaton, A.F. (2011), Aggregating semantic concepts for event representation in lifelogging, *Proceedings of the International Workshop on Semantic Web Information Management*, pp. 8:1–8:6.
25. Wang, P., Yang, S.Q., Sun, L.F., Smeaton, A.F. (2016). What are the Limits to Time Series Recognition of Semantic Concepts?, *International Conference on Multimedia Modeling 2016*, Miami, Fl, USA.
26. Wang, P., Smeaton, A.F. (2012). Semantics-based selection of everyday concepts in visual lifelogging. *International Journal of Multimedia Information Retrieval*, 1(2). pp. 87-101.
27. Hamm J. , Stone, B., Belkin, M., Dennis, S. (2012). Automatic Annotation of Daily Activity from Smartphone-Based Multisensory Streams, *4th International Conference, MobiCASE 2012*, pp 328-342.
28. Dobbins C., Merabti M., Fergus, P., Jones, D., Bouhafs, F. (2013). Exploiting linked data to create rich human digital memories, *Computer Communications* 36 (2013) 1639–1656
29. Li, Y., Guo, Y.K. (2016). Wiki-Health: From Quantified Self to Self-Understanding, *Future Generation Computer Systems*, Volume 56, 2016, Pages 333–359
30. Wang, P., Smeaton, A.F., Mileo, A.. Semantically Enhancing Multimedia Lifelog Events, *15th Pacific-Rim Conference on Multimedia*, 1-4 Dec, 2014, Kuching, Malaysia.

Demonstrations List

(D1) A low cost, user friendly system for the collection and analysis of ultrasonic sound with synchronous video.
Schenk, Katrin; Wojciechowski, Noelle; Desgain, Zach

(D2) The Virtual Foodscape Simulator – gaming, designing and measuring food behaviour in created food realities.
Mikkelsen, Bent Egberg

(D3) BehavioMatrix: A Behaviour Measuring System.
McIntyre, Gordon James; Göcke, Roland

(D4) Wireless shirt for ECG and Respiration.
Dimov, Alexandar

(D5) xobs: a free open-source video analysis programme.
Sullivan, Matthew Stephen

(D6) Integrated eye tracking and behavioral coding.
Richard Lilley

(D7) Integrated solutions for simulator research.
van Mil, Bram; Theuws, Hans

(D8) Consumer Friendly Virtual Reality Simulation for Randomized Controlled Trials in Consumer Research.
van den Puttelaar, Jos; Verain, Muriel; Onwezen, Marleen

(D9) Innovative Food Intake App based on 2 Hour Recalls.
van den Puttelaar, Jos; Verain, Muriel; Onwezen, Marleen

(D10) FaceReader: new developments in facial expression analysis.
Theuws, Hans; den Uyl, Tim

(D11) A fully integrated operant 'box' for adult zebrafish.
Budenberg, William James; Brennan, Caroline H; Parker, Matthew; Brock, Alistair James

Author Index

MEASURING BEHAVIOR 2016

DUBLIN - IRELAND

25 - 27 MAY 2016

www.measuringbehavior.org