Noldus

The Observer XT in HCI research

Tobias Heffelaar

MB 2010 Eindhoven, Netherlands, 24 August 2010

Noldus

Scope of this course

- Theme: Measuring user-system interaction
- Context: Research or testing
- For: Academia, industry or consulting

Noldus

Measuring user-system interaction

From usability…

- Measuring effectiveness, efficiency, satisfaction
- In relation to well-defined tasks
- Primarily for productivity tools, office applications

…to user experience

- Measuring the complete user experience
- Emotion, fun, excitement, trust
- For consumer products, games, e-learning systems, web sites

Noldus

User-Centered Design

Focus on the user: observe and listen

- Ethnography - field studies
- Focus groups
- Formative usability testing
  - Low-fidelity prototypes
  - High-fidelity prototypes
- Summative usability testing

Professional tools make this work pleasant, efficient and effective!

Noldus

Course objectives

What you will learn

- learn how to design, execute and analyze observational studies (such as usability tests)
- learn how to configure and use The Observer in various stages of the research and development process, from field observations to summative usability tests
- learn how to combine observational data collection with other techniques, such as eye tracking and physiological data acquisition

Noldus

Course agenda

Designing the coding scheme

Data collection

Multimodal measurements

Analyzing observational data

HILAS case study

Noldus

Your instructor

Tobias Heffelaar, Usability specialist (at)

Noldus Information Technology

Wageningen
The Netherlands

Leesburg
Virginia, USA

www.noldus.com/usability

Noldus

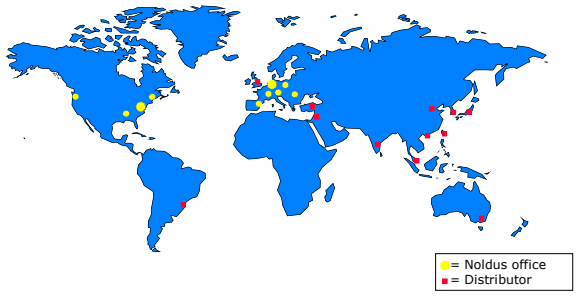
Company profile

Who we are
- Provider of software, hardware and services for HCI research and usability testing
- Founded in 1989
- Currently ~90 employees

What we do
- Development of data collection and analysis software (installed base: ~4,000 organizations in >75 countries)
- System integration (mobile and stationary usability labs)
- Lab rental
- Training
- Consulting

Noldus

International Offices



Noldus

Sources of data and illustrations

Tools manufacturers
- Biopac
- Noldus Information Technology
- Polar
- SensoMotoric Instruments
- TechSmith
- Tobii Technology
- VicarVision

Usability practitioners
- Many collaborators and clients

Noldus

Course agenda

Designing the coding scheme
- From basic to elaborate, from generic to specific, best practices
- Freeform annotation vs. structured event logging
- Optimizing the coding scheme to make scoring easier

Data collection
- Real-time logging vs Post-test annotation

Design a simple and effective coding scheme for both qualitative and quantitative data collection

Noldus

Free-form annotation

Write down what is said and observed during test
- Quick, no equipment required
- Difficult to analyze, expert observer required, how fast can you write?

Audio recording of participant comments, post-hoc transcription
- Easy, nothing gets lost
- Context is gone, analysis is laborious, subjective data collection

Noldus

Free-form annotation

User comments
…….
Uhhh…. What do I do here?
Now I see what this is!
So I just click here…(user clicks button)
Oh no (application closes)
Where is my data? Did it save my file?
(restarts application) I hope my document is still there
….

Noldus

Structured event logging

Pro
- Pre-defined set of events
- Pre-defined usability goals
- More objective data collection
- Quantitative analysis possible
- Allows validation and reliability analysis
- Facilitates education and training of observers

Con
- Time-consuming
- Requires preparation
- Specify what to code (not always known beforehand)
- Train coding scheme

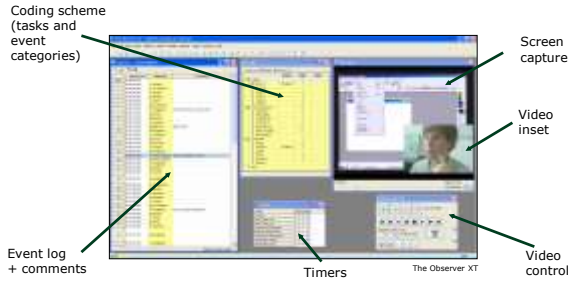
Noldus

Best of both worlds

- Start with free-form annotation
- Build coding scheme as you go along

Noldus

Structured event logging



Noldus

Methods compared

	Free-form annotation	**Structured event logging and coding**
Data produced	Qualitative	Quantitative
Timing accuracy	Not so important	Very important
When used	Formative tests with few participants	Summative tests with many participants, human factors research
Data used for	Communication with client, compilation of video highlights	Usability metrics, sequential analysis, pattern detection

Noldus

Structured data collection

Logging of important events, such as:

Tasks
- Keep track of user intention

Interface elements
- Menu items, commands, controls, …

Participant behavior
- Comments, gestures, facial expression, …

Encountered problems
- Navigation issues, feedback issues, …

Problem solving strategies
- Trial-and-error, use of info resources, …

Noldus

Coding scheme elements

- Name (subject, behavior, modifier)
- Hierarchical level (grouping)
- Definition
- Shortcut key
- State or event
- Sound feedback (e.g. spoken element)
- Image (e.g. facial expression)
- Video clip (e.g. gesture)

Noldus

Specific coding scheme

Example: interaction between user and ATM

User action	System action
Enter card	Process input
Enter code	Eject card
Specify amount	Offer money
Take money	Wait
Take card	
Wait	

Noldus

Simple coding scheme

Example: logging user remarks and problems encountered while working on a task

Task	Action	Communication
Task 1	Problem	Comment
Task 2	Error	Question
Task 3	Usability hit	Vocalization (sigh, etc.)
Task 4		
Task 5		
Task 6		

Noldus

Scoring events

- Record events and time of occurrence
- Live or offline: depending on level of detail needed

- Use mouse to select events from list, or
- Type shortcut keys: no need to look away from subject

Noldus

Hierarchical coding scheme

- Model tasks or task structures
- Code at different levels (inheritance)
- Permits fine-grained data analysis

The Observer XT

Noldus

Observing multiple participants

Why?

- Usability testing of systems for two or more multiple concurrent users
- Analyzing focus groups

Noldus

Focus group coding scheme

Person
Moderator
Participant 1
Participant 2
Participant 3

Verbal communication
Remark
Question
Explanation

Non-verbal communication
Gesture
Eye contact
Looking away
Vocalization (sigh, etc.)

Noldus

Multi-user applications

	Same time	**Different times**
Same place	Control rooms, meeting rooms	Project scheduling, coordination tools
Different places	Video conferencing, instant messenging, multi-player games	Email, bulletin boards

Noldus

Coding two participants

The Observer XT

- Screen real-estate becomes limiting factor
- Requires large screen or dual-display setup

Noldus

Multi-user usability tests

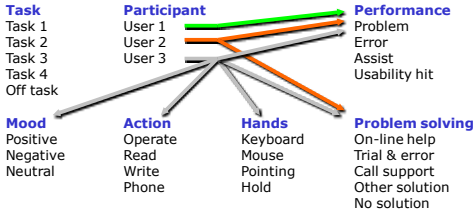
Methodological and technical challenges

- Coding issues: record individual behaviors as well as interactions between participants, real-time data entry requires multiple loggers
- Group behavior is complex: roles can change, different interaction patterns
- Experiments are more difficult to control: selection of subject groups, designing experiment and tasks
- Technical issues: video recording, data gathering, large amount of data for analysis

Noldus

Complex multi-user coding scheme

- Code events for each user
- Code certain events for certain users



Noldus

Complex coding scheme

Task intention coding (partial)
(Dzida et al.)

Prepare
Object to be treated
Suitable tools
Goal to be defined

Execute
Command
Data input

Mood
Positive
Negative
Neutral

Opportunism
Adapt result
Ad-hoc search
Prevent damage

Feedback
Offer interaction element
Visual guidance
Offer context
Present result
Prompt user

System action
Alert
Response

User evaluation
Check result
Stop
Comment
Retry
Stress

Noldus

Typical MMI evaluation coding scheme

Task list
Task 1
Task 2
Etc

Tasks are states (durations)→ time on task

Usability
System error
User error
Assist
Use of support (or OLH, etc)

All elements are point events→ number of occurences

Communication
Positive comment
Negative comment
Positive non-verbal expression
Negative non-verbal expression

Noldus

Tools for multi-user tests

- Multiple video cameras, microphones and/or 'live' observers
- Multiple mixing and recording equipment
- Hardware for recording and playback of multiple synchronized video and audio streams
- Event logging software that supports multiple loggers and/or subjects

Noldus

Observational data collection
Things to keep in mind

- Structured data collection takes more time before and during test, but it speeds up analysis and improves the quality of the results
- When observing interactions between users: coding scheme extra important
- Event logging equipment exists for just about any test

Noldus

Course agenda

Multimodal measurements

- Eye tracking, physiological data acquisition, keystroke and mouse logging
- Combining The Observer with other data collection tools: synchronization of measurements, data import
- Integrated visualization of video streams, events and signals

Noldus

Multimodal data collection

Measuring multiple modalities of user-system interaction

Behavioral

- Task performance
- Keyboard activity
- Mouse activity
- Body posture
- Facial expression
- Eye movement
- Gestures
- Verbal comments

Physiological

Emotional state

- Galvanic skin resistance

Mental load

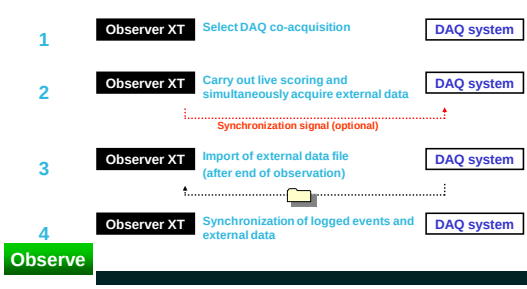
- Pupil diameter
- Heart rate variability
- Respiration

Physical load

- Electromyogram
- Grip force

Noldus

Working with external data
basic steps



Noldus

Why do I need synchronization?

- **Synchronization enables you to examine physiological data in relation to the associated logged events**

Observe

Noldus

Methods of synchronization

- **Automatic synchronization**
  - **XT computer and DAQ system connected through the synchronization signal**
- **Manual synchronization**
  - **No connection between XT computer and DAQ system**

Observe

Noldus

Visualize Data

Create a chart

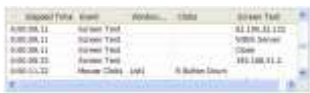
Analyze

Noldus

Automatic event logging (1)

- Log mouse clicks, keystrokes, URLs, dialogs, pop-up menus, etc.
- X,Y coordinates logged

- **Pro**: easy and cost-effective data acquisition
- **Con**: hard to interpret unless combined with observational data and task information



Morae

Noldus

Automatic event logging (2)

Caution
- Keyboard and mouse logging tools can create massive amounts of data
- To reduce data volume: select in advance what to track
- Available tools vary in logging capabilities and 'intelligence'

uLog

Noldus

Automated observation with sensors

contact pressure

ambient light

loudness

motion

proximity

vibration

Photos: Infusion Systems Ltd

Noldus

Measuring eye gaze

- Tool: eye tracking system, head-mounted or contact-free
- Measures:
  - where the subject looks
  - how long and often they look at something
  - path the eyes follow between predefined areas of interest
  - pupil diameter (measure of cognitive load)
- Complement observational methods, e.g. find co-occurrence of "confusion" and "fixate"

Tobii Technology

SensoMotoric Instruments

Noldus

Eye tracking: analysis

Gaze plot

Hot spots

Noldus

Physiological measurements

- Heart rate variability, skin conductance, blood pressure: indicator of stress level
- Integrate physiology with video, observational data, eye movement, system events…

Noldus

Physiological measurements

The Observer XT

Noldus

Usability testing of aviation systems

Air Traffic Control simulator

Cockpit

Data sources: video, audio, task performance, EOG, EPOG, ECG, EEG, Ear Pulse, etc.

Noldus

Facial Electromyography (EMG)

- EMG measures contraction of muscles
- Electrodes applied to skin surface
- Can be used to measure difficulty of motor tasks, e.g. UI operations: cursor positioning, drag & drop, resizing windows
- But: rather laborious and obtrusive

Photo: Ab de Haan, Radboud University, Nijmegen, The Netherlands

Photo courtesy of Richard Hazlett

Noldus

Other signals

Finger clip
- Skin temperature
- Skin conductance
- Pulse frequency
- Pulse amplitude

Objective assessment of feeling (stress, fear, anger, relaxation, concentration)

BioMedical

TIM-Lab, Danube University Krems, Krems, Austria

Noldus

Facial expression recognition

- Automatic classification of facial expressions and emotions
- Ingredients:
  - Facial Action Coding System (Paul Ekman)
  - Active Appearance Modeling (Cootes & Taylor)
  - Real-time image analysis software
  - Fast computer
- First commercial products becoming available:
  - six basic emotions classified automatically

Vicar Vision

Noldus

Measuring user behavior in virtual environments

- Usability testing of hardware designs
- Usability testing of games

Create VR environment

Log user actions

Review user behavior in sync with VR display

Analyze performance

Noldus

Multimodal measurements
Things to keep in mind

- **Eye tracking and physiological measurements can greatly enrich HCI data**
- **Measurements can be rather invasive**
- **Technical skills needed for data acquisition**
- **Integration and synchronization with video streams and logged events is not trivial**

Noldus

Calculate Statistics
Statistics available

For Behavior Analysis:

- **Minimum duration**
- **Maximum duration**
- **Total duration**
- **Total number**
- **Mean duration**
- **Standard deviation of duration**
- **Standard error of duration**
- **Rate per minute**
- **Percentage (of observation)**
- **Percentage (of interval)**

Analyze

Noldus

Select Data
Definitions

Grouping

means that two or more observations or elements of the coding scheme are analyzed as one entity

Filtering

means that you choose a subset of elements to be displayed or used in quantitative analyses

Nesting

means that you analyze time segments based on an event (or a combination of events) scored for one or more subjects

Analyze

Noldus

Course agenda

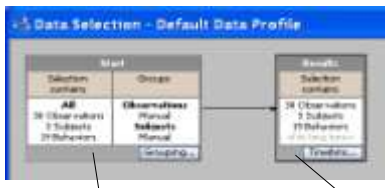
Analyzing observational data

- Qualitative analysis: reviewing tests, creating video highlights
- Quantitative analysis: computing performance measures and usability metrics

Noldus

Select Data

Default data profile



Analyze

Noldus

Select Data

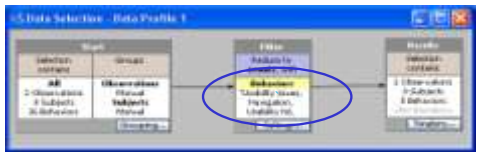
Create your own data selection

Analyze

Noldus

Select Data
Filtering

Use Filter to choose the observations, subjects and event types to analyze

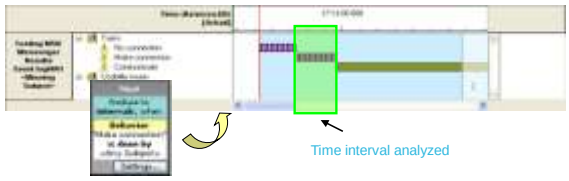


Analyze

Noldus

Select Data
Nesting

The consequences of Nesting over data

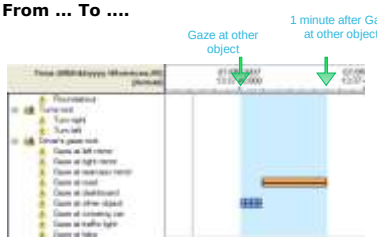


Analyze

Noldus

Select Data
Free intervals

Use the Free Intervals to analyze the time From … To ….



Analyze

Noldus

Reliability Analysis

Checking the validity of your scores

- Intra-observer reliability — For checking your own consistency
- Inter-observer reliability — For training observers

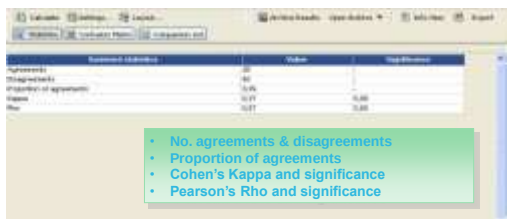
Noldus

Reliability Analysis

- Frequency / Sequence based method used
- Each event in both data files are considered
- The Observer XT searches for matching events within a defined tolerance window
- The total number of agreements and disagreements are calculated

Noldus

Reliability Analysis
Result - Statistics



- No. agreements & disagreements
- Proportion of agreements
- Cohen's Kappa and significance
- Pearson's Rho and significance

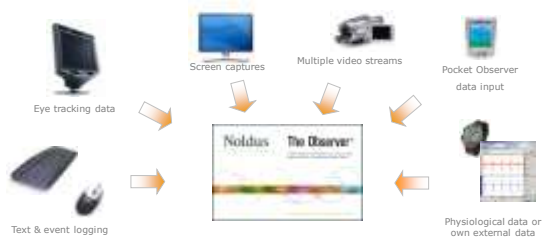
Noldus

HILAS: a HF case study

Noldus

Noldus

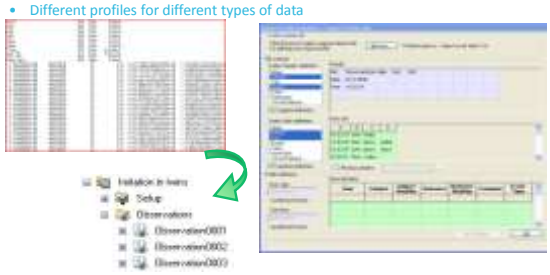
Importing data



Noldus

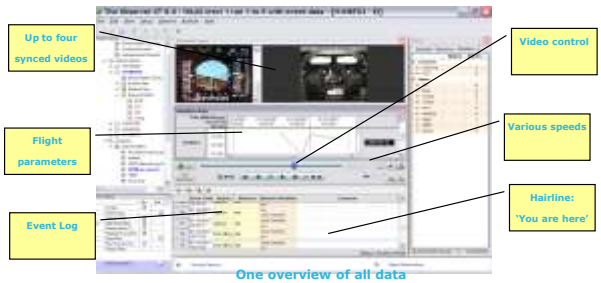
Import profiles

- Specification of the way data files should be intepreted
- Different profiles for different types of data



Noldus

Integrated Visualization



One overview of all data
Accurate and visual quality assessment

Noldus

Calculate Statistics

Behavior Analysis

- Example: The total amount of time a pilot communicates with the control tower

Numerical Modifier Analysis

- Example: The average variation in plane altitude when flying manual

Lag Sequential Analysis

- Example: The most common action taken when a problem is detected

Reliability Analysis

- Example: The consistency of observations between observers

Noldus

The experts

2 expert raters marked all SA relevant events using:
'SA positive' or
'SA negative' events

Unfortunately no correlation between markings

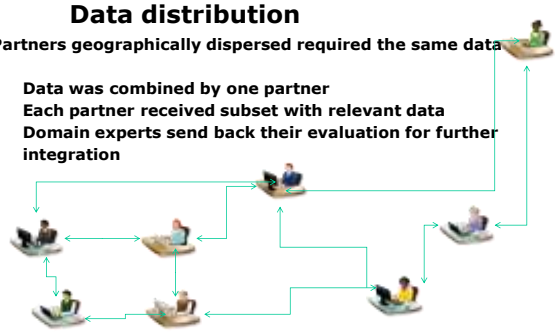
General impression (more pos o
per observation) did match.

Noldus

Data distribution

Partners geographically dispersed required the same data

- **Data was combined by one partner**
- **Each partner received subset with relevant data**
- **Domain experts send back their evaluation for further integration**



Noldus

Independent Variables

Keep track of Trial and/or Crew characteristics
Used when filtering data for analysis

Noldus

The Observer XT

Manually marking significant events
Integrating multiple data sources
Analyzing events
Filtering of data
Export to statistical programs

Noldus

HILAS measures

Measuring multiple modalities

Behavior
Situation awareness
Communication between pilots and ATC

Simulator data
Flight simulator data (Continuous)
Pilot comments (e.g. Frequency changes)

Physiology

Mental load
Heart rate (variability)
Pupil diameter, dwell time, blink rate

Physical load
Grip force
Facial temperature

Noldus

Thank you for your attention!

Please don't forget to fill out the evaluation form