Applying Nonlinear Dynamics Features for Speech-based Fatigue Detection

Jarek Krajewski[1]
krajewsk@uni-wuppertal.de

David Sommer[2]
d.sommer@fh-sm.de

Thomas Schnupp[2]
t.schnupp@fh-sm.de

Tom Laufenberg[1]
laufenberg@uni-wuppertal.de

Christian Heinze[2]
c.heinze@fh-sm.de

Martin Golz[2]
m.golz@fh-sm.de

[1]University Wuppertal
Experimental Industrial Psychology
Gaußstraße 20, 42097 Wuppertal, Germany

[2]University of Applied Sciences Schmalkalden
Neuroinformatics and Signal Processing
Blechhammer, 98574 Schmalkalden, Germany

ABSTRACT

This paper describes a speech signal processing method to measure fatigue from speech. The advantages of this realtime approach are that obtaining speech data is non obtrusive, free from sensor application and calibration efforts. Applying methods of Non Linear Dynamics (NLD) provides additional information regarding the dynamics and structure of fatigue speech comparing to the commonly applied speech emotion recognition feature set (e.g. fundamental frequency, intensity, pause patterns, formants, cepstral coefficients). We achieved significant correlations between fatigue and NLD features of 0.29. The validity of this approach is briefly discussed by summarizing the empirical results of a sleep deprivation study.

Author Keywords

Speech signal processing, nonlinear dynamics features, speech emotion recognition, human-computer-interaction.

ACM Classification Keywords

H.5.5 Sound and music computing

INTRODUCTION

The prediction and warning of drivers against impending critical fatigue play an important role in preventing accidents and the resulting human and financial costs. Hence, many efforts have been reported in the literature for measuring fatigue related states [1-5]. In contrast to electrode- or video-based instruments, the utilization of voice communication as an indicator for fatigue could match the demands of everyday life measurement. Contact free measurements as voice analysis are non-obtrusive (not interfering with the primary driving task) and favorable for fatigue detection since an application of sensors would cause annoyance, additional stress and often impairs working capabilities and mobility demands. In addition, speech is easy to record even under extreme environmental conditions (bright light, high humidity and temperature), requires merely cheap, durable, and maintenance free sensors and most importantly, it utilizes already existing communication system hardware. Furthermore, speech data is omnipresent in many professional driver settings. Nevertheless, several sources of noise during driving, as e.g. motor sound, radio, and sidetalk can lead to difficult recording situations. In sum, given these obvious advantages, the renewed interest in computational demanding analyses of vocal expressions has been enabled just recently by the advances in computer processing speed.

An important aspect in the vocal tract during fatigue influenced speech production is the generation of nonlinear aerodynamic phenomena including non-laminar flow, flow separation in various regions, generation and propagation of vortices and formation of jets rather than well-behaved laminar flow [6-8]. The collapse of laminar flow arises at high reynolds number. Due to the relevant length and subsonic speed of air flow in the vocal tract, this number is very large, indicating that the air flow can be expected to be turbulent. The air jet flowing through the vocal tract during speech production includes convoluted paths of rapidly varying velocity, which are highly unstable and oscillate between its walls, attaching or detaching itself, and thereby changing the effective cross-sectional areas and air masses.

Several issues are responsible for the generation of these nonlinear effects: The vocal folds behave as a vibrating valve, disrupting the constant airflow from the lungs and forming it into regular puffs of air. Modeling approaches which have their origin in fluid dynamics coupled with the elastodynamics of a deformable solid understand this phonation process as nonlinear oscillation: dynamical forcing from the lungs provides the energy needed to overcome dissipation in the vocal fold tissue and vocal tract air. The vocal folds themselves are modeled as elastic tissue with nonlinear stress-strain relationship. These nonlinear stretching qualities of the vocal folds are based on larynx muscles and cartilage which produces nonlinear behavior. Furthermore, vocal tract and the vocal folds are coupled when the glottis is open resulting in significant changes in formant characteristics between open and closed glottis cycles. The movement of the vocal folds themselves is modeled by a lumped two mass system connected by springs again with nonlinear coupling. These nonlinear phenomena produce turbulent flow while the air jet may be modulated either by the vibration of the walls or by the generated vortices. Several methods based on chaotic dynamics and fractal theory have been suggested to describe these aerodynamic turbulence related phenomena of the speech production system [9-17] including the modeling of the geometrical structures in turbulence (spatial structure, energy cascade) utilizing fractals and multifractals [15-19], nonlinear oscillator models [20-22], and state-space reconstruction. This state-space reconstruction is done utilizing the embedding theorem which reconstructs a multidimensional attractor by embedding the scalar signal into a phase space. The embedding allows us to reconstruct the geometrical structure of the original attractor of the system which formed the observed speech signal. Moreover, it helps us to discover the degree of determinism of an apparently random signal, e.g. by applying measures like Lyapunov exponents.

However, no empirical research has been done to examine the turbulence effects in speech signals, which might be induced by fatigue related change of heat conduction within the vocal tract. Previous work associating changes in voice with fatigue [23-24] has generally focused only on features derived from speech emotion recognition [25] whereas nonlinear dynamics based speech features [26] have received no attention. Thus, the aim of this study is to apply nonlinear dynamics (NLD) based features within the field of speech acoustics in order to improve the prediction of fatigue.

METHODS

Seventeen students, recruited from the University of Wuppertal (Germany), took part in this study voluntarily. The participants were instructed to maintain their normal sleep pattern and behaviour. We conducted a within-subject sleep deprivation design (8.00 p.m to 4.00 a.m). During the night of sleep deprivation a well-established, standardized self-report fatigue measure, the Stanford Sleepiness Scale (SSS), was used by the subjects and two experimental assistants just before the recordings to determine the fatigue state. On this scale, a score of 1 point indicates “feeling active and vital, alert, wide awake” and a score of 7 points indicates “almost in reverie, sleep onset soon, losing struggle to remain awake” (M = 5.06; SD = 2.02). During the night, the subjects were confined to the laboratory and supervised throughout the whole period. Between sessions, they remained in a room, watched DVD, and talked. Non caffeinated beverages and snacks were available ad libitum.

Speech Material and Recording. The recording took place in a laboratory room with dampened acoustics using a high-quality, clip-on microphone (sampling rate: 44.1 kHz, 16 bit). The input level of the sound recording was kept constant throughout the recordings. Furthermore the subjects were given sufficient prior practice so that they were not uncomfortable with this procedure. The verbal material consisted of a long vowel [o:] extracted from a German phrase: “Rufen Sie den N[o:]tdienst” (“Please call the ambulance”). The sentence was taken from simulated communication with a driver assistance system. The participants recorded other verbal material at the same session, but in this article we focus on the material described above (17 subjects x 4 sentences = 68 speech samples).

Seventeen students, recruited from the University of Wuppertal (Germany), took part in this study voluntarily. The participants were instructed to maintain their normal sleep pattern and behaviour. We conducted a within-subject sleep deprivation design (8.00 p.m to 4.00 a.m). During the night of sleep deprivation a well-established, standardized self-report fatigue measure, the Stanford Sleepiness Scale (SSS), was used by the subjects and two experimental assistants just before the recordings to determine the fatigue state. On this scale, a score of 1 point indicates “feeling active and vital, alert, wide awake” and a score of 7 points indicates “almost in reverie, sleep onset soon, losing struggle to remain awake” (M = 5.06; SD = 2.02). During the night, the subjects were confined to the laboratory and supervised throughout the whole period. Between sessions, they remained in a room, watched DVD, and talked. Non caffeinated beverages and snacks were available ad libitum.

Speech Material and Recording. The recording took place in a laboratory room with dampened acoustics using a high-quality, clip-on microphone (sampling rate: 44.1 kHz, 16 bit). The input level of the sound recording was kept constant throughout the recordings. Furthermore the subjects were given sufficient prior practice so that they were not uncomfortable with this procedure. The verbal material consisted of a long vowel [o:] extracted from a German phrase: “Rufen Sie den N[o:]tdienst” (“Please call the ambulance”). The sentence was taken from simulated communication with a driver assistance system. The participants recorded other verbal material at the same session, but in this article we focus on the material described above (17 subjects x 4 sentences = 68 speech samples).

Feature Extraction. The following feature families within nonlinear time series analysis are computed: (a) *State space features (375#).* To extract the nonlinear properties of the speech angle signal, we computed a three-dimensional state space (sound pressure, Δ sound pressure, $\Delta\Delta$ sound pressure), and a reconstructed phase space ($d = 3,4$; $\tau = 1$;

sound pressure t_0, sound pressure t_1, sound pressure t_2). The geometrical properties of the resulting attractor figures were described by trajectory based descriptor contours (angle between consecutive trajectory parts, distance to centroid of attractor, length of trajectory leg). The temporal information of the contours was captured by computing functionals; (b) *Fractal features (110#).* They try to quantify self-affinity and underlying complexity of the speech signal, e.g. box counting dimension, Cao's minimum embedding dimensions, correlation dimension, fractal dimension, information dimension; (c) *Entropy features (5#).* They assess the regularity/irregularity or randomness of speech signal fluctuations, as e.g. the Lyapunov exponents, which characterize the system's "degree of chaos" by estimating the exponential rate of divergence or convergence of nearby orbits on its phase-space. Positive Lyapunov exponents indicate divergence of nearby orbits and thus long-term unpredictability.

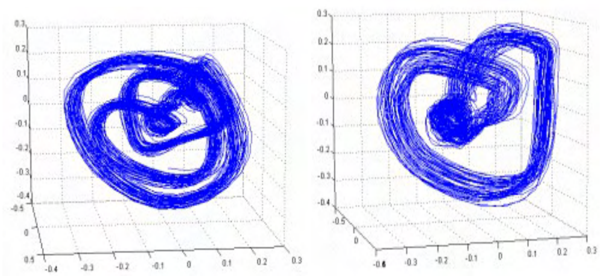

Figure 1. Reconstructed phase spaces (d = 3 ; τ = 10; sound pressure t_0, sound pressure t_1, sound pressure t_2) for an alert (left) and fatigue (right) speech sample [a:]. The geometrical properties of the resulting attractor figures were described e.g. by trajectory based descriptor contours (angle between consecutive trajectory parts, distance to centroid of attractor, length of trajectory leg) or entropy features as Lyapunov exponents, which characterize the system's "degree of chaos" by estimating the exponential rate of divergence or convergence of nearby orbits.

RESULTS

In order to quantify the association between fatigue and (a) commonly applied SER speech features and (b) NLD features, we computed Pearson's correlation coefficients The significant correlations are displayed in Table 1 and 2.

SER Features	r
standard deviation of intensity	-.29*
range of fundamental frequency	-.24*
jitter (small fundamental frequency perturbations)	-.33*
mean of delta-mel-frequency cepstrum coefficient 7	.34*
mean of formant 3 position (third maximum of vocal tract transfer function; resonance frequency)	.29*

Table 1. Correlation of SER Features and fatigue; * p < .05.

NLD Features	r
skewness of vector length within state space = (sound pressure, Δ sound pressure, ΔΔ sound pressure)	-.21*
skewness of vector length within reconstructed phase space = (τ = 1, d = 3)	-.22*
mean of Cao's minimum embedding dimensions = (τ = 1, d_{max} = 3, number nearest neighbors = 3)	.25*
mean of Cao's minimum embedding dimensions = (τ = 8, d_{max} = 4, number nearest neighbors = 3);	.25*
mean of Cao's minimum embedding dimensions = (τ = 10, d_{max} = 4, number nearest neighbors = 3)	.26*
mean of Cao's minimum embedding dimensions (τ = 8, d_{max} = 5, number nearest neighbors = 3);	.27*
mean of Cao's minimum embedding dimensions (τ = 10, d_{max} = 5, number nearest neighbors = 3).	.29*

Table 2. Correlation of NLD Features and fatigue; * p < .05.

DISCUSSION

Due to several nonlinear phenomena producing turbulent air flow applying NLD speech feature might provide additional information regarding the dynamics and structure of sleepy speech comparing to the commonly applied SER feature set. We achieved significant correlations between SSS and the NLD feature 'Cao's minimum embedding dimension' of .29, which indicates a higher complexity of fatigue speech samples. Explanations for this effect might be found in the following effect chain: fatigue induced decreased body temperature → reduced heat conduction within the vocal tract, changed friction between vocal tract walls and air, changed laminar flows, jet streams, and turbulences → higher complexity of a more turbulent speech signal.

Several factors might have influenced the results obtained by NLD methods and consequently have to be considered, e.g. recording duration, degree of stationarity, and superimposed noise. Furthermore, it would seem advisable that future studies address the main topics of enriching the NLD feature set with further fractal (multifractal analysis, power-law correlation, detrended fluctuation analysis), entropy (approximate entropy/sample entropy, multiscale entropy, compression entropy), symbolic dynamics measures, and delay-vector-variance [1,27]. Additionally, it would seem beneficial that future studies address the main topics of enriching the steering feature set with easy accessible driving related informations as e.g. pedal movement behaviour. Finally, it has to be emphasized that up to now a reference measure of fatigue (gold standard of fatigue) is lacking. But approximating fatigue by concordance of self and several observer ratings might act as an intermediate solution, that follows the general procedure of finding ground truth values within the speech emotion recognition research community. In conclusion, methods derived from NLD could offer promising insights into fatigue induced speech changes. They supply additional information and complement traditional time- and frequency-domain analyses of speech.

REFERENCES

1. Sommer, D. Analyse des Mikroschlafs mit Methoden der Computergestützten Intelligenz. Aachen: Shaker, 2009.
2. Sommer D, Golz M, Schupp T, Krajewski J, Trutschel U, & Edwards D A measure of strong driver fatigue. *Proceeding International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, *4* (2009) 9-15.
3. Schnupp T, Schenka A, Edwards D, Krajewski J, & Golz M Is posturography a candidate for a vigilance test? World Congress on Medical Physics and Biomedical Engineering, *IFMBE Proceedings*, *25*, 9 (2009), 388-392.
4. Krajewski J, Sommer D, Trutschel U, Edwards D, & Golz M. Steering wheel behavior based estimating of fatigue. *Proceeding International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, *4* (2009), 118-124.
5. Altmüller T Driver Monitoring and Drowsiness Detection by Steering Signal Analysis. Aachen: Shaker, 2007.
6. Teager, HM & Teager, SM Evidence for nonlinear sound production mechanisms in the vocal tract. Speech Production and Speech Modelling, W. J. Hardcastle and A. Marchal, Eds., vol. 55 of NATO Advanced Study Institute Series D, Bonas, France (1989), 241–261.
7. Thomas, TJ A finite element model of fluid flow in the vocal tract. *Comput. Speech & Language*, *1*, (1986), 131–151.
8. Kaiser, J.F. Some observations on vocal tract operation from a fluid flow point of view. In I. R. Titze and R. C. Scherer, editors, Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control, Denver Center for Performing Arts, Denver, CO (1983), 358–386.
9. Banbrook, M., McLaughlin, S, & Mann, I Speech characterization and synthesis by nonlinear methods. IEEE Trans. *Speech Audio Processing*, *7* (1999), 1–17.
10. Dimitriadis, D. & Maragos, P. Robust energy demodulation based on continuous models with application to speech recognition. In Proc. Eurospeech-03, Geneva, Switzerland, 2003.
11. Dimitriadis, D, Katsamanis, N, Maragos, P, Papandreou, G & Pitsikalis, V Towards automatic speech recognition in adverse environments. Proceedings of 7th Hellenic European Conference on Research on Computer Mathematics and Its Applications (HERCMA '05), Athens, Greece, 2005.
12. Maragos, P, Kaiser, JF, & Quatieri, TF Energy separation in signal modulations with application to speech analysis. *IEEE Trans.on Signal Processing*, *41*, 10 (1993), 3024–3051.
13. Narayanan, S. & Alwan, A. A nonlinear dynamical systems analysis of fricative consonants. *J. Acoust. Soc. Am.*, *97*, 4 (1995), 2511–2524.
14. McLaughlin, S & Maragos, P Nonlinear methods for speech analysis and synthesis. In Marshall S, Sicuranza G. Hindawi (Eds.) Advances in nonlinear signal and image processing, Volume 6, 2007.
15. Maragos, P & Potamianos, A Fractal dimensions of speech sounds: Computation and application to automatic speech recognition. *J. Acoust. Soc. Am.*, *105*, 3 (1999), 1925–1932.
16. Maragos, P. Fractal aspects of speech signals: Dimension and interpolation. Proc. IEEE ICASSP-91, 1991.
17. Ashkenazy, Y The use of generalized information dimension in measuring fractal dimension of time series. *Physica A*, *271* 3-4 (1999), 427–447.
18. Adeyemi, O & Boudreaux-Bartels, FG Improved accuracy in the singularity spectrum of multifractal chaotic time series. In Proc. IEEE, ICASSP-97, Munich, Germany, 1997.
19. Ashkenazy, Y The use of generalized information dimension in measuring fractal dimension of time series. *Physica A*, *271*, 3-4 (1999), 427–447.
20. Kubin, G Synthesis and Coding of Continuous Speech with the Nonlinear Oscillator Model. In Proc. ICASSP'96, vol. 1, Atlanta, USA, 1996, 267.
21. Quatieri, TF, & Hofstetter, EM Short-time signal representation by nonlinear difference equations. In Proc. IEEE , ICASSP'90, Albuquerque, 1990.
22. Townshend, B Nonlinear prediction of speech signals. IEEE Trans. Acoust., Speech, Signal Processing, 1990.
23. Krajewski J, Batliner A, & Golz M Acoustic sleepiness detection – Framework and validation of a speech adapted pattern recognition approach. *Behavior Research Methods*, *41* (2009), 795-804.
24. Krajewski J, Batliner A, & Wieland R Multiple classifier applied on predicting microsleep from speech. In ICPR (Eds.), 19th Conference on Pattern Recognition (4 pages, no pagination). Tampa/ Florida: IEEE Computer Society Press, 2009.
25. Batliner, A, Seppi, D, Steidl, S & Schuller, B Segmenting into adequate units for automatic recognition of emotion-related episodes: a speech-based approach. *Advances in Human-Computer Interaction* (in press).
26. Banbrook, M, McLaughlin, S. & Mann, I Speech characterization and synthesis by nonlinear methods. *IEEE Trans. Speech Audio Processing*, *7*, (1999), 1–17.
27. Voss A, Schulz S, Schroeder R, Baumert M, & Caminal P Methods derived from nonlinear dynamics for analysing heart rate variability. *Philos Transact A Math Phys Eng Sci.*, *367*, 1887 (2009), 277-296.