Fatigue in Air Traffic Communication – Combining Acoustic Features within an Computational Intelligence Approach

Jarek Krajewski¹, Rainer Wieland¹, David Sommer², Martin Golz²

¹University of Wuppertal, Germany, Work and Organizational Psychology
krajewsk@uni-wuppertal.de; wieland@uni-wuppertal.de

²Univ. of Applied Sciences Schmalkalden Germany, Neuro Computer Science, and Signal Processing
d.sommer@fh-sm.de; m.golz@fh-sm.de

Abstract. A promising approach for the real-time detection of sleepiness in Air Traffic Controller (ATC) and pilots is the acoustic sleepiness analysis. This article describes a general framework for detecting accident-related fatigue states based on prosody, articulation and speech quality related speech characteristics. The advantages of this realtime measurement approach are that obtaining speech data is non obstrusive, and free from sensor application and calibration efforts. The main part of the feature computation is the combination of frame level based speech features and high level contour descriptors resulting in over 22,064 features per speech sample. In general the measurement process follows the speech adapted steps of pattern recognition: (a) recording speech, (b) preprocessing (segmenting speech units of interest), (c) feature computation (using perceptual and signal processing related features, as e.g. fundamental frequency, intensity, pause patterns, formants, cepstral coefficients), (d) dimensionality reduction, (e) classification (Support Vector Machine), and (f) validation (10-fold cross validation). The validity of this approach is briefly discussed by summarizing the empirical results of a sleep deprivation study (N=12; 01.00-08.00 a.m.). Our results are limited mainly by the fact that the findings were based on noise-free recordings.

Keywords: Acoustic, Computational Intelligence, Fatigue, Accident Prevention, Signal Processing

1 Detecting Fatigue for Air Traffic Safety

It is a commonly accepted fact that sleepiness causes cognitive impairments including the area of vigilance, working memory, short term memory, executive function, mathematical processing, cognitive speed, spatial orientation, motor control supervisory control (Durmer & Dinges, 2005), problem solving, decision making, situation awareness, divergent thinking capacity, and verbal creativity (Kollias, Amir, Kim & Grandjean, 2004; Nilsson, Soderstrom, Karlsson et al., 2005; Rogers, Dorrian & Dinges, 2005; Wesensten, Killgore & Balkin, 2006). Furthermore sleepiness related cognitive impairments have been recognized as a critical
factor for a broad range of road traffic accidents (MacLean, 2003). Similar to truck driving, in aviation long-haul flights are characterized by extended time-on-task durations and disrupted circadian rhythms, which both could lead to dangerous levels of fatigue and sleepiness (Wright & McGown, 2001). Accordingly 21% of the reported incidents mentioned in the Aviation Safety Reporting System (including pilots and Air Traffic Controller, ATC) were related to fatigue (Roske-Hofstrand, 1995). The measurement of fatigue and sleepiness is thus an important concern for air traffic safety and other safety sensitive fields as e.g. aeronautics, chemical factories and nuclear power stations (Melamed, 2002).

Hence, many efforts have been reported in the literature for developing measures for sleepiness and microsleep events (e.g. Sommer, Chen, Golz, Trunschel & Mandic, 2005). These systems mainly focus on information such as (a) eye blinking, eyelid movement, and saccade eye movement (Caffier, 2002; Galley, 2007), (b) EEG data (Golz, Sommer & Mandic, 2006) as well as (b) gross body movement, head movement, mannerism, and facial expression in order to characterize the state of alertness. Apart from these promising advances in analysing facial and gestural expressivity, there has been recently renewed interest in vocal expression and speech analysis. Mainly this fact is promoted by the progress in speech science. Using voice communication as an indicator of sleepiness would have the following advantages: obtaining speech data is non-invasive, annoyance free, non obstructive, and most important it is omnipresent in aviation. Air traffic control (ATC) has relied on the voice radio for communication between aircraft pilots and air traffic control operators since its beginning. Given the aeronautical life cycle constraints, it is expected that the analogue radio will remain in use well beyond 2020.

In this paper we describe an acoustic measurement approach in order to measure fatigue states. Our attention is focused particularly on the processing step of feature extraction. The rest of this paper is organized as follows: In Section 2 computing high level contour descriptor features is explained. Section 3 describes the validation of the acoustic measurement approach using a simulator driven sleep deprivation study. A brief discussion of the results, limitations and future work is given in Section 4.

2 Acoustic Features

The following fatigue related physiological changes can influence voice characteristics: (a) decreased muscle tension (reduced facial expression and smiling, unconstricted pharynx, softening of vocal tract walls, vocal fold elasticity and tension), (b) decreased body temperature (reduced heat conduction, changed viscoelasticity of vocal folds, changed friction between vocal tract walls and air as well as impaired laminar flows), (c) reduced cognitive processing speed (impaired speech planning) and (d) flat and slow respiration (low subglottal pressure). The corresponding auditive-perceptual effects might be lower pitch, loudness, articulatory precision, and rate of articulation, as well as shift in the voice quality to a breathier voice. Finding quantitative measures (acoustic features) for these perceptual changes and combining them within methods of computational intelligence and pattern recognition could lead to the successful identification of fatigue in speech.
Acoustic features can be divided according to auditive-perceptual concepts into prosody (pitch, intensity, rhythm, pause pattern, speech rate), articulation (slurred speech, reduction and elision phenomena), and speech quality (breathy, tense, sharp, hoarse, modal voice) related features (Schuller, Batliner, Seppi et al., 2007; Batliner, Steidl, Schuller et al. 2006). Another distinction can be drawn from using signal processing categories as time, frequency or phase space domain features. Our approach prefers the fusion of perceptual features and purely signal processing and speech recognition based features without any known auditive-perceptual correlates. Typical frame level based acoustic features (Low-Level Descriptors, LLD; see Vlasenko, Schuller, Wendemuth & Rigoll, 2007) used in emotion speech recognition and audio processing (Mierswa & Morik, 2005) are fundamental frequency (acoustic correlate to pitch; maximum of the autocorrelation function), intensity, duration of voiced/unvoiced segments, harmonics-to-noise ratio, position and bandwidth of 6 formants (resonance frequencies of the vocal tract depending strongly on its actual shape), 15 linear predictive coding coefficients, 12 mel frequency cepstrum coefficients (“spectrum of the spectrum”), and 12 linear frequency cepstrum coefficients (without the perceptually oriented transformation into the mel frequency scale). Tab 1.

After splitting the speech signal into frames and computing the above mentioned frame level features, the values of each frame level feature are connected to contours. This procedure results in 56 speech feature contours (e.g. the fundamental frequency contour, the bandwidth of formant 4 contour etc.), which are joined by their first and second derivates (velocity (∆) and acceleration (∆∆) contours). Furthermore these 168 speech feature contours are described by elementary statistics (linear moments, values and positions of extrema, quartiles, ranges, length of time periods beyond threshold values, regression coefficients, etc.), and spectral descriptors (spectral energy of low frequency bands vs. high frequency bands, etc.) resulting in 22,064 high-level speech features (56 speech contours x 394 functionals), see Fig 1.

Fig. 1. Processing flow of acoustic feature computation including the computation of frame level based features and contour descriptors (functionals) to capture sufficient temporal information.

3 Empirical Validation Results

We conducted a within-subject sleep deprivation design (N = 12; 01.00 - 08.00 a.m). During the night of sleep deprivation a well established, standardised self-report sleepiness measure, the Karolinska Sleepiness Scale (KSS) was used by the subjects and 2 experimental assistants nearly every hour just before the speech recordings. The verbal material consisted
of a simulated pilot-air traffic controller communication ("Cessna nine three four five Lima, County tower, runway two four in use, enter traffic pattern, report left base, wind calm, altimeter three zero point zero eight"). The participants recorded other verbal material at the same session, but in this article we focus on sustained phonation only. For training and classification purposes, the records were further divided in two classes: sleepy (SS) and non sleepy (NSS) with the microsleep validated boundary value $KSS \geq 7.5$ (8 samples per subject; total number of speech samples: 95 samples; 34 samples NSS, 61 samples SS; $KSS:=\text{mean of the three KSS-Ratings}; M=7.49; SD=2.28$). During the night, the subjects were confined to the laboratory, conducting a driving simulator task and were supervised throughout the whole period.

![Graph showing average fundamental frequency (pitch) contours of air traffic communication for sleepy vs. alert speakers.](image)

**Fig. 2.** Average fundamental frequency (pitch) contours of air traffic communication for sleepy vs. alert speakers.

The best three single features have the following correlations to self-reported sleepiness: 'speaker normalized - fundamental frequency - mean delta peaklocation' = .56; 'speaker normalized - intensity value of 90$^{th}$ Percentil' = -.53; 'linear frequency cepstrum coefficient 07-time beyond threshold of median + 2*median'. In order to determine the multivariate prediction
performance a multiple linear regression was conducted. The prediction on unseen test data (10-fold cross validation) reached a multiple correlation of $R = .51$ (mean absolute prediction error = 1.87). The recognition rate (number of cases classified correctly divided by all cases; RR) of a second multivariate method (Support Vector Machine; sigmoid kernel function; $C = .4$; gamma = 0.000001) was 72.3%, the class-wised averaged classification rate (mean of recognition rates for each class; CL) was 71.9%.

4 Concluding Remarks

The main findings of the present study may be summarised as follows. First, acoustic features extracted from simulated pilot-air traffic controller communication contain information about sleepiness states. Secondly, in our experiments on a two-class classification problem (sleepy vs. non sleepy speech), we achieved an accuracy rate of over 99% on unseen data with a SVM classifier. Thirdly the analyzing of a single phrase is sufficient for the detection of critical sleepiness states. Due to the hypothesized sleepiness related physiological changes in cognitive speech planning, respiration, phonation, articulation, and radiation, the results for the reported classification performance above chance level were largely as could be expected. This is consistent with previous sleepiness related findings that suggest an association of acoustic features (Krajewski & Kroeger, 2007) with sleepiness.

There are some limitations of this study. First, the air-ground voice communication between the air traffic controller and the aircraft is done over an analogue VHF radio channel. There are several factors which can influence the speech signal quality (Hofbauer & Kubin, 2006) as the multipath propagation (depending on the geometric dimensions - e.g. surface slope - and the properties of the objects in a scene – e.g. conductivity, dielectric constant, roughness), the Doppler Effect (depending on the speed and angle of arrival of of the aircraft), the channel attenuation (path losses depending on the distance and the obstacles between transmitter and receiver) and the additive noise (thermal noise in electronic components, atmospheric noise, radio channel interference, or engine ignition noise). Due to these characteristics of air-ground radio communication it still remains unclear to which degree the detection of sleepy pilots will be impaired.

References


Caffier, P.P. (2002). The Spontaneous Eye-Blink as Sleepiness Indicator in Patients with Obstructive Sleep Apnoea Syndrome-a Pilot Study. Sleep Medicine, 2, 155-162.


