INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Cross-Domain Classification of Drowsiness in Speech: The Case of Alcohol Intoxication and Sleep Deprivation

Yue Zhang[1], *Felix Weninger*[2], *Björn W. Schuller*[1]

[1]Department of Computing, Imperial College London, London, U.K.
[2]Nuance Communications, Ulm, Germany

yue.zhang1@imperial.ac.uk, felix@weninger.de, bjoern.schuller@imperial.ac.uk

Abstract

In this work, we study the drowsy state of a speaker, induced by alcohol intoxication or sleep deprivation. In particular, we investigate the coherence between the two pivotal causes of drowsiness, as featured in the Intoxication and Sleepiness tasks of the INTERSPEECH Speaker State Challenge. In this way, we aim to exploit the interrelations between these different, yet highly correlated speaker states, which need to be reliably recognised in safety and security critical environments. To this end, we perform cross-domain classification of alcohol intoxication and sleepiness, thus leveraging the acoustic similarities of these speech phenomena for transfer learning. Further, we conducted in-depth feature analysis to quantitatively assess the task relatedness and to determine the most relevant features for both tasks. To test our methods in realistic contexts, we use the Alcohol Language Corpus and the Sleepy Language Corpus containing in total 60 hours of genuine intoxicated and sleepy speech. In the result, cross-domain classification combined with feature selection yields up to 60.3 % unweighted average recall, which is significantly above-chance (50 %) and highly notable given the mismatch in the training and validation data. Finally, we show that an effective, general drowsiness classifier can be obtained by aggregating the training data from both domains.

Index Terms: Computational Paralinguistics, speaker states, drowsiness detection, transfer learning, feature analysis

1. Introduction

Drowsiness exhibits a distinct pattern of effects on human behaviour, including sleepiness, lethargy, and latency, as well as reduced control and coordination of speech articulation, phonation and respiration [1]. The severity of the drowsiness-related cognitive, physiological impairment can be observed in many instances. Drowsy driving due to alcohol intoxication or sleep deprivation frequently results in fatal traffic accidents and is identified on police reports as one main driver risk factor [2]. Given the nature of the medium-term speaker states continuing over a certain period, people are usually unaware of the onset of drowsiness, e. g., after alcohol intake. For this reason, automatic monitoring and warn systems which detect diminishing vigilance and performance impairment are of vital importance [3]. Other possible applications relate to the digital health care domain, forensics, and work contexts, especially in safety sensitive areas such as chemical factories, nuclear power stations, and air traffic control.

Drowsiness – which can be induced by inebriation, fatigue, and health issues – influences the way a person speaks. Perceptual studies have shown that a person listening to voice samples can reliably discriminate between sober and intoxicated speech according to various linguistic and phonetic properties [4, 5, 6]. Hence, much research effort has been undertaken to analyse the measurable characteristics indicating the effects of alcohol on human speech, including acoustic features [7, 8, 9, 10], articulatory characteristics [11, 12, 13], and emotional aspects [14]. In particular, the bulk of studies found that the rate of speech and the overall amplitude decrease after alcohol consumption, whereas the pitch variability, the mean fundamental frequency (F0), and the sentence duration increase [1]. Likewise, sleepy speech manifests changes in prosody (e. g., monotonic intonation, shifted speech rate, reduced syllable duration due to retarded cognitive speech planning), articulation (e. g., slurred pronunciation, speech errors, disfluency), and voice quality (e. g., nasal or breathy speech) [15]. Specifically, in accordance with the findings of alcohol intoxication, the sleepiness-induced changes of speech parameters include a general decrease in speech rate [16] and an increased average absolute deviation of intensity [17]. On the other hand, quite opposite effects, for instance a decreased mean fundamental frequency (F0), have been observed [18, 19].

Aside from the contributions within the INTERSPEECH Speaker State Challenge [20], only a few studies reported on attempts to detect alcohol intoxication and sleepiness from the speech signal by means of statistical classification [21, 22, 17]. The work [23] first pointed out the synergistic and antagonistic effects on a high level when combining sleep deprivation and moderate alcohol consumption. However, to the authors' best knowledge, the acoustic correlates *between* alcohol intoxication and sleepiness have never been regarded so far, despite their common effects in relation to drowsiness. In this work, we aim to shed light on the coherence between these two speaker states by means of cross-domain experiments and acoustic feature analysis, This approach is especially meaningful given the scarcity of genuine intoxicated speech data since usually plenty of speech is available from speakers being sober, but rarely when being inebriated. Finally, due to the fact that the actual cause of drowsiness is often unknown in practical situations, we suggest to jointly using the data from both domains to train a general drowsiness classifier.

2. Data sets

In this section, we briefly describe the collection of the drowsy speech data considered for our study. More detailed descriptions can be found in [15].

2.1. Alcohol Speech Corpus (ALC)

The Alcohol Language Corpus (ALC) [24, 25] comprises 162 speakers (84 male, 78 female) within the age range 21–75, mean age 31.0 years and standard deviation 9.5 years, from 5 different locations in Germany. For our experiments, the same gender balanced subset is used as in the Speaker State Challenge (154

http://dx.doi.org/10.21437/Interspeech.2017-1015

Table 1: *Partitioning into train, development, and test set according to the setup of the Speaker State Challenge. 'NAL/AL' and 'NSL/SL' denote recordings of non-alcoholised/ alcoholised and non-sleepy/ sleepy speakers*

(a) ALC

#	**NAL**	**AL**	Σ
Train	3 750	1 650	5 400
Develop	2 790	1 170	3 960
Test	1 620	1 380	3 000
Σ	8 160	4 200	12 360

(b) SLC

#	**NSL**	**SL**	Σ
Train	2 125	1 241	3 366
Develop	1 836	1 079	2 915
Test	1 957	851	2 808
Σ	5 918	3 171	9 089

speakers, 77 male/female). Speakers voluntarily underwent a systematic intoxication test supervised by the staff of the Institute of Legal Medicine, Munich. Before the test, each speaker chose the blood alcohol concentration (BAC) he or she wanted to reach during the intoxication test. The BAC was measured for each subject 20 minutes after alcohol intake, ranging between 0.28 and 1.75 per mill. Immediately after the BAC test, each speaker was asked to perform the ALC speech test which lasted no longer than 15 minutes, to avoid significant changes in BAC. The corpus further contains control recordings from each speaker in sober condition, which took place in the same acoustic environment. Three different speech styles are part of the ALC speech test: read speech, spontaneous speech, and command & control.

The recordings from the intoxication experiment were assigned to the classes non-alcoholised ('NAL') and alcoholised ('AL') based on the threshold of 0.5 per mill BAC value, corresponding to the legal limit for driving in Germany and other countries. The sober recordings were associated with the 'NAL' class, accordingly. The recordings are randomly partitioned into speaker-independent and gender balanced data sets (cf. Table 1).

2.2. Sleepy Language Corpus (SLC)

The Sleepy Language Corpus (SLC) contains recordings from 99 participants (56 female, 43 male), who took part in six partial sleep deprivation studies. The age range of the participants is 20–52 years, with a mean of 24.9 and a standard deviation of 4.2 years. The speech material consists of different tasks including sustained vowel phonation, read speech, and spontaneous speech. A well established, subjective sleepiness questionnaire based on the Karolinska Sleepiness Scale (KSS) was filled by the subjects (self-assessment) and additionally by two experimental assistants from a perceptive point of view.

The scores ranging from 1 (extremly alert) to 10 (cannot stay awake) were discretised into not sleepy ('NSL') and sleepy ('SL') at the threshold of 7.5. This threshold (between level 7 'sleepy, some effort to stay awake' and level 8, 'very sleepy, great effort to stay awake') marks the occurrence of microsleep events, leading to a significant increase in accident risk [26]. For automatic classification, a speaker-independent partitioning was used that is additionally stratified by gender and study setup (environment and degree of sleep deprivation). The distribution of instances is given in Table 1.

3. Methodology

3.1. Feature extraction

The *ComParE* set of supra-segmental acoustic features is a well-evolved set for automatic recognition of paralinguistic speech phenomena, as used for the baseline of the INTERSPEECH ComParE series. It contains 6 373 static features resulting from the computation of various functionals over low-level descriptor (LLD) contours. The configuration file is provided in the public release of openSMILE [27, 28]. Important subgroups of the ComParE feature set comprise prosodic (*PROS*), Mel Frequency Cepstral Coefficients (*MFCC*), spectral (*SPEC*), and voice quality (*VQ*) features, which represent highly relevant acoustic attributes as mentioned in Section 1.

3.2. Classification

For transparency and reproducibility, we use open-source implementations of Support Vector Machines (SVMs) from the WEKA data mining toolkit [29]. As training algorithm, we use Sequential Minimal Optimisation (SMO) [30]. In cross-domain experiments, feature standardisation is crucial to alleviate covariate shift. To this end, we follow a straightforward scheme that does not require batch processing at test time. The training set of each corpus (ALC, SLC) is standardised to zero mean and unit variance. When evaluating on a development set, it is standardised using the scales and offsets computed on the training set of the same database. When evaluating on the test set, the scales and offsets are computed on the union of training and development set of the same corpus. No further standardisation is applied during training and testing. Due to the skewed class distribution in both the ALC and SLC, the instances of the minority class (NAL and NSL) in the data used for training the classifier are duplicated once, so as to reach an approximately balanced class distribution while guaranteeing reproducibility.

3.3. Feature selection

We found feature selection to be crucial for the performance for cross-domain experiments. There, we strive to identify features which carry similar meaning with respect to two different causes of drowsiness in speech. In this section, we detail the methods we evaluated in our study.

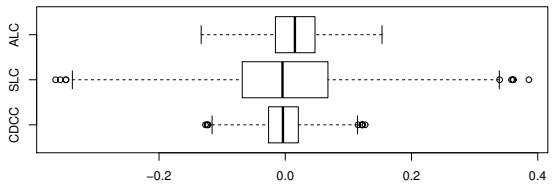
3.3.1. Correlation-based feature selection

A widely used method to reduce the feature space dimension is correlation-based feature selection (CFS) [29], which we use as a baseline method. This technique aims to minimise inter-feature correlations while maximising correlation with the class labels. There, the merit M of a feature subset S with k features is given by

$$M(S) = \frac{k\,\mathrm{CC}_{\mathrm{cf}}}{\sqrt{k + k(k-1)\mathrm{CC}_{\mathrm{ff}}}}, \qquad (1)$$

where $\mathrm{CC}_{\mathrm{cf}}$ denotes the mean correlation coefficient (CC) of features in S with the class label, and $\mathrm{CC}_{\mathrm{ff}}$ is the average CC of features in S with each other. It is easy to see that a candidate subset that maximises M will provide an optimal trade-off between high predictive power regarding the class label (numerator) and low redundancy among the features (denominator). To apply CFS for identifying suitable cross-domain features, we simply measure the merit on the union of the training sets of both domains.

Figure 1: *Box-and-whisker plots of the distribution of CCs of the 6 373 ComParE acoustic features on the ALC and SLC data sets, as well as the corresponding CDCCs.*



3.3.2. Selection by cross-domain correlation coefficient

We also apply the cross-domain correlation coefficient (CDCC) measure introduced in the work [31], which is particularly tailored to cross-domain experiments. The purpose of the CDCC measure is to weigh high correlation in single domains against correlation deviations across different domains. Denoting by i and j the domains (here: alcohol intoxication and sleepiness), the definition of CDCC is

$$\mathrm{CDCC}^2_{f,i,j} = \frac{\left|r_f^{(i)} + r_f^{(j)}\right| - \left|r_f^{(i)} - r_f^{(j)}\right|}{2}, \qquad (2)$$

where $r_f^{(i)}$ is the correlation of feature f with the domain i. It is obvious that the CDCC measure is symmetric in the sense that $\mathrm{CDCC}^2_{f,i,j} = \mathrm{CDCC}^2_{f,j,i}$, and that it ranges from -1 to 1. If a feature f exhibits either strong positive or strong negative correlation with both domains, the CDCC will be near one, whereas it will be near -1 if a feature is strongly positively correlated with one domain yet strongly negatively correlated with the other. A CDCC near zero indicates that the feature is not significantly correlated with both domains (although it might still be correlated with either one). Thus, we can expect similar performance across domains when selecting features with high CDCC. In our study, we cross-validated the usage of the top 50, 100, or 200 features.

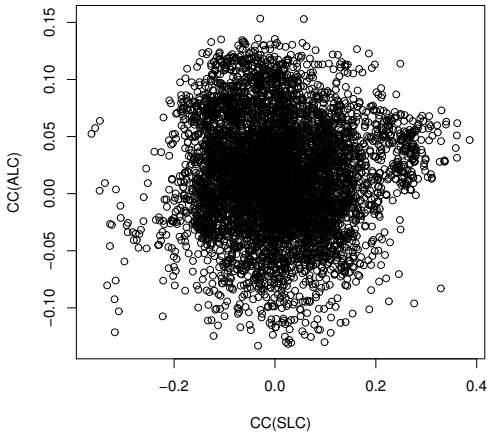
4. Experiments and Results

4.1. Feature analysis

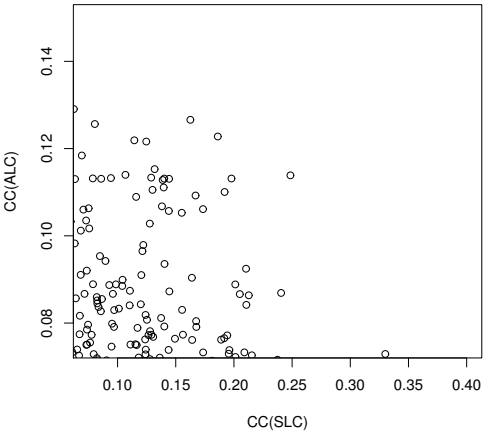
In a first experiment, we assess the predictive power of the acoustic features for within-domain and cross-domain intoxication and sleepiness classification, we compute the point-biserial correlation coefficient (CC) of each acoustic feature with the target label (alcohol intoxication/ sleepiness), as well as the CDCC. The chosen form of CC is equivalent to the Pearson CC with the nominal target label converted to a numeric label with values 0 (non-alcoholised/ non-sleepy) and 1 (alcoholised/ sleepy). The distribution of the three coefficients across the acoustic features is shown in Figure 1. It can be seen that the correlations on the ALC are much lower than on the SLC, but the value range of the CDCC resembles the one on the ALC. This is promising insofar as previous results [15] indicate that reasonably robust classification is possible on the ALC corpus despite the generally low feature-label correlations.

Next, to shed light on the cross-domain consistency of the predictive power of the acoustic features, and hence the resemblance of the tasks of intoxication and sleepiness classification from speech, we plot the CCs on the ALC and SLC training sets in Figure 2. It can be seen that measured across all features, there is no strong coherence of the tasks in the acoustic space[1]. However, there do exist features that are positively correlated with both intoxication and sleepiness[2]. This fact motivates the usage of CDCC-based feature selection.

Figure 2: *Scatterplots of the point-biserial correlation coefficients (CC) of ComParE acoustic features with the target labels of the ALC and SLC training sets.*



(a) All features (6 373)



(b) 122 features with minimum CC of .075 on each task

4.2. Classification results

In the following, we present the cross-domain classification results, i.e., training on (non-)alcoholised and testing on (non-)sleepy speakers, and vice versa. We compare these with the in-domain baseline, i. e., training and testing on disjoint subsets of the the same corpus (ALC or SLC), as was the task of the Speaker State Challenge [20]. Following the Challenge procedure, the evaluation measure is unweighted average recall (UAR). The complexity constant C, which is one of the most important hyper-parameters of SVM training, is chosen from the set $10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}$ according to the highest UAR on the development set. Furthermore, since the number of se-

[1] It has to be noted, though, that the correlation of the CCs on ALC and SLC ($\rho = -.0270$) is significant at the 5 % level according to a two-tailed t-test.

[2] For space constraints, we do not show a scatterplot of features that are negatively correlated with both tasks – these do exist as well.

Figure 3: *Intoxication classification results by percentage of unweighted average recall (UAR) on the ALC development set when training on the SLC and using 50, 100, or 200 features selected by best CDCC.*

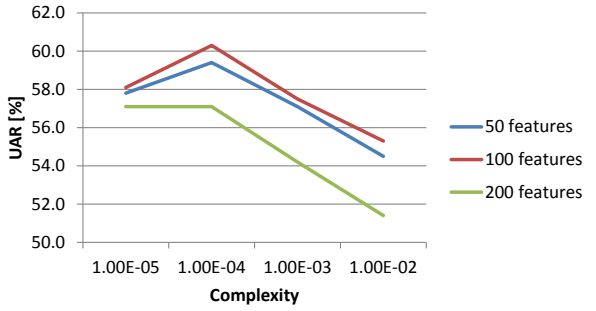
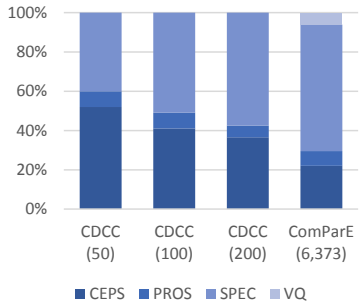


Figure 4: *Breakdown of feature sets into percentages of CEPStral, PROSodic, SPECtral and Voice Quality features. CDCC(n): Feature selection of n features by highest CDCC.*



lected features is the primary hyper-parameter of the CDCC selection method, we tune it in a preliminary experiment on the development set of the ALC. The results of the latter are depicted in Figure 3. By comparison, 100 features provide the best performance, which is consistent with our observation that only a limited number of features show 'reasonable' correlation on both tasks (cf. Figure 2). Breaking down the selected features by CDCC into feature groups (cf. Figure 4), we observe a generally increasing importance of cepstral features when restricting ourselves to using less and less features, while the relative importance of prosodic features remains constant. Voice quality features are not selected at all by the CDCC method. Table 2 shows the results of the in-domain baselines and cross-domain classification in terms of UAR. Note that the baseline results are slightly better than those reported in the Speaker State Challenge, which can be attributed to using the ComParE acoustic feature set, which is somewhat larger and more carefully designed than the one used for the Challenge.

As a rule of thumb derived from a one-tailed z-test (at the 5 % level) [32], UARs above 52.2 % can be considered significantly above chance (50 %), which holds for both the ALC and SLC test sets due to their similar size.

Summarising the results, we note that on the development set of the ALC, 60.3 % UAR can be obtained by training on the SLC training set and using CDCC-based feature selection. This significantly outperforms the result obtained by CFS. Using the entire feature set, only chance level results can be obtained. The picture is similar on the ALC test set, but overall performance is lower – still, the CDCC result is significantly above chance. On

Table 2: *In-domain and cross-domain classification results by percentage of unweighted average recall (UAR). Feature selection by correlation-based feature selection (CFS) and cross-domain correlation coefficient (CDCC, top 100 features). C: Complexity constant for SVM training, tuned by UAR on the development set of ALC/SLC.*

(a) ALC, training on the ALC (in-domain), SLC (cross-domain), or joint training

Train on / UAR	C	Dev	Test
ALC	10^{-4}	66.1	66.5
SLC	10^{-5}	50.1	47.7
SLC (CFS)	10^{-5}	54.7	51.4
SLC (CDCC)	10^{-4}	60.3	55.4
ALC+SLC	10^{-4}	65.5	63.3

(b) SLC, training on the SLC (in-domain), ALC (cross-domain), or joint training

Train on / UAR	C	Dev	Test
SLC	10^{-4}	66.5	72.0
ALC	10^{-5}	47.3	50.2
ALC (CFS)	10^{-3}	50.7	59.5
ALC (CDCC)	10^{-2}	51.6	54.7
ALC+SLC	10^{-4}	62.7	69.0

the SLC development set, none of the cross-domain classification methods yields a result above chance, yet on the SLC test set the CFS-based selection gives 59.5 % UAR when training on the ALC training and development set.

Finally, investigating joint training with both the ALC and SLC, we observe worse classification performance on the test sets than when training on only in-domain data. Nevertheless, these results are promising in the light of a general classification of drowsiness where the potential cause (intoxication or sleep deprivation) is unknown at test time. Averaging across the ALC and SLC test sets, data aggregation yields 66.2 % UAR, compared to 63.0 % / 63.7 % (training only on ALC or SLC, test on both), without requiring knowledge of the reason of drowsiness.

5. Conclusion

In this study, we have provided evidence for the acoustic coherence of two drowsiness related phenomena, namely alcohol intoxication and sleepiness, in speech. While the overall correlation of acoustic descriptors with the target labels severely differs between the two, the proposed usage of cross-domain feature selection by CFS or CDCC enabled classification accuracies significantly above chance on the ALC and SLC development and test sets, as well as the SLC test set, with only out-of-domain training data. These results are all the more notable as the ALC and SLC data comprise real-world recordings of spontaneous as well as command-and-control speech, and only a fraction of the baseline features (100 vs. 6.3 k) are used. Furthermore, robust classification of general drowsiness can be achieved by joint training. Future work will exploit multi-task learning such as by means of deep neural networks with shared hidden layers in order to model task interrelations in a common feature space [33].

6. Acknowledgments

The research work has received funding from the EU Horizon 2020 Framework Programme with the Research & Innovation Action under grant agreement No. 645378 (ARIA-VALUSPA).

7. References

[1] O. M. Cooney, K. McGuigan, P. Murphy, and R. Conroy, "Acoustic analysis of the effects of alcohol on the human voice," *Journal of the Acoustical Society of America*, vol. 103, no. 5, p. 2895, 1998.

[2] J. C. Stutts, J. W. Wilkins, J. S. Osberg, and B. V. Vaughn, "Driver risk factors for sleep-related crashes," *Accident Analysis & Prevention*, vol. 35, no. 3, pp. 321–331, 2003.

[3] T. A. Dingus, H. L. Hardee, and W. W. Wierwille, "Development of models for on-board detection of driver impairment," *Accident Analysis & Prevention*, vol. 19, no. 4, pp. 271–283, 1987.

[4] D. B. Pisoni and C. S. Martin, "Effects of alcohol on the acoustic-phonetic properties of speech: Perceptual and acoustic analyses," *Alcoholism: Clinical and Experimental Research*, vol. 13, no. 4, pp. 577–587, 1989.

[5] K. Johnson, D. B. Pisoni, and R. H. Bernacki, "Do voice recordings reveal whether a person is intoxicated? A case study," *Phonetica*, vol. 47, no. 3–4, pp. 215–237, 1990.

[6] F. Schiel, "Perception of alcoholic intoxication in speech," in *Proc. of INTERSPEECH*, (Florence, Italy), pp. 3281–3284, IEEE, 2011.

[7] F. Klingholz, R. Penning, and E. Liebhardt, "Recognition of low-level alcohol intoxication from speech signal," *Journal of the Acoustical Society of America*, vol. 84, no. 3, pp. 929–935, 1988.

[8] H. Hollien, G. De Jong, C. A. Martin, R. Schwartz, and K. Liljegren, "Effects of ethanol intoxication on speech suprasegmentals," *Journal of the Acoustical Society of America*, vol. 110, no. 6, pp. 3198–3206, 2001.

[9] A. Braun, H. J. Künzel, D. Recasens, J. Romero, and M. Solé, "The effect of alcohol on speech prosody," in *Proc. of the International Congress of Phonetic Sciences*, vol. 2645, (Barcelona, Spain), pp. 2645–2648, International Phonetic Association, 2003.

[10] F. Schiel, C. Heinrich, and V. Neumeyer, "Rhythm and formant features for automatic alcohol detection," in *Proc. of INTERSPEECH*, (Makuhari, Japan), pp. 458–461, ISCA, 2010.

[11] F. Trojan and K. Kryspin-Exner, "The decay of articulation under the influence of alcohol and paraldehyde," *Folia Phoniatrica et Logopaedica*, vol. 20, no. 4, pp. 217–238, 1968.

[12] L. C. Sobell, M. B. Sobell, and R. F. Coleman, "Alcohol-inducted disfluency in non-alcoholics," *Folia Phoniatrica et Logopaedica*, vol. 34, no. 6, pp. 316–323, 1982.

[13] D. M. Behne, S. M. Rivera, and D. B. Pisoni, "Effects of alcohol on speech: I. durations of isolated words, sentences, and passages in fluent speech," *The Journal of the Acoustical Society of America*, vol. 90, no. 4, pp. 2311–2311, 1991.

[14] J. J. Curtin, C. J. Patrick, A. R. Lang, J. T. Cacioppo, and N. Birbaumer, "Alcohol affects emotion through cognition," *Psychological Science*, vol. 12, no. 6, pp. 527–531, 2001.

[15] B. Schuller, S. Steidl, A. Batliner, F. Schiel, J. Krajewski, F. Weninger, and F. Eyben, "Medium-term speaker states – a review on intoxication, sleepiness and the first challenge," *Computer Speech and Language, Special Issue on Broadening the View on Speaker Analysis*, vol. 28, no. 2, pp. 346–374, 2014.

[16] A. P. Vogel, J. Fletcher, and P. Maruff, "Acoustic analysis of the effects of sustained wakefulness on speech," *The Journal of the Acoustical Society of America*, vol. 128, no. 6, pp. 3747–3756, 2010.

[17] J. Krajewski, A. Batliner, and M. Golz, "Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach," *Behavior Research Methods*, vol. 41, no. 3, pp. 795–804, 2009.

[18] B. Johannes, V. P. Salnitski, H.-C. Gunga, and K. Kirsch, "Voice stress monitoring in space–possibilities and limits," *Aviation, Space, and Environmental Medicine*, vol. 71, no. 9, pp. A58–65, 2000.

[19] T. L. Nwe, H. Li, and M. Dong, "Analysis and detection of speech under sleep deprivation," in *Proc. of INTERSPEECH*, (Pittsburgh, PA), pp. 1846–1849, ISCA, 2006.

[20] B. Schuller, A. Batliner, S. Steidl, F. Schiel, and J. Krajewski, "The INTERSPEECH 2011 Speaker State Challenge," in *Proc. of INTERSPEECH*, (Florence, Italy), pp. 3201–3204, ISCA, 2011.

[21] M. Levit, R. Huber, A. Batliner, and E. Noeth, "Use of prosodic speech characteristics for automated detection of alcohol intoxication," in *Proc. of Workshop on Prosody and Speech Recognition*, (Red Bank, NJ), pp. 103–106, ISCA, 2001.

[22] M. Sigmund, A. Prokes, and P. Zelinka, "Detection of alcohol in speech signal using lf model," in *Proc. of International Conference on Artificial Intelligence and Applications*, (Innsbruck, Austria), pp. 193–196, ACTA Press, 2010.

[23] D. Headley, "The combined effects of alcohol and sleep deprivation on cognitive functioning: a review," *Alc. Techical Reports*, vol. 5, pp. 45–51, 1976.

[24] F. Schiel and C. Heinrich, "Laying the foundation for in-car alcohol detection by speech," in *Proc. of INTERSPEECH*, (Brighton, U.K.), pp. 983–986, ISCA, 2009.

[25] F. Schiel, C. Heinrich, and S. Barfüsser, "Alcohol Language Corpus," *Language Resources and Evaluation*, vol. 46, no. 3, pp. 503–521, 2011.

[26] M. Ingre, T. Åkerstedt, B. Peters, A. Anund, G. Kecklund, and A. Pickles, "Subjective sleepiness and accident risk avoiding the ecological fallacy," *Journal of Sleep Research*, vol. 15, no. 2, pp. 142–148, 2006.

[27] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor," in *Proc. of ACM Multimedia*, (Florence, Italy), pp. 1459–1462, ACM, 2010.

[28] F. Eyben, F. Weninger, F. Groß, and B. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in *Proc. of ACM Multimedia*, (Barcelona, Spain), pp. 835–838, ACM, 2013.

[29] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[30] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[31] F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro, and K. R. Scherer, "On the acoustics of emotion in audio: What speech, music and sound have in common," *Frontiers in Psychology, section Emotion Science, Special Issue on Expression of emotion in music and vocal communication*, vol. 4, no. Article ID 292, pp. 1–12, 2013.

[32] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, 1998.

[33] Y. Zhang, Y. Liu, F. Weninger, and B. Schuller, "Multi-Task Deep Neural Network with Shared Hidden Layers: Breaking Down the Wall between Emotion Representations," in *Proc. of ICASSP*, (New Orleans, LA), pp. 4990–4994, IEEE, 2017.