Snoring - An Acoustic Definition

Christoph Janott[1], Christian Rohrmeier[2], Maximilian Schmitt[3], Werner Hemmert[1], and Björn Schuller[3,4]

Abstract— **OBJECTIVE The distinction of snoring and loud breathing is often subjective and lies in the ear of the beholder. The aim of this study is to identify and assess acoustic features with a high suitability to distinguish these two classes of sound, in order to facilitate an objective definition of snoring based on acoustic parameters.**

METHODS A corpus of snore and breath sounds from 23 subjects has been used that were classified by 25 human raters. Using the openSMILE feature extractor, 6 373 acoustic features have been evaluated for their selectivity comparing SVM classification, logistic regression, and the recall of each single feature.

RESULTS Most selective single features were several statistical functionals of the first and second mel frequency spectrum-generated perceptual linear predictive (PLP) cepstral coefficient with an unweighted average recall (UAR) of up to 93.8%. The best performing feature sets were low level descriptors (LLDs), derivatives and statistical functionals based on fast Fourier transformation (FFT), with a UAR of 93.0%, and on the summed mel frequency spectrum-generated PLP cepstral coefficients, with a UAR of 92.2% using SVM classification. Compared to SVM classification, logistic regression did not show considerable differences in classification performance.

CONCLUSION It could be shown that snoring and loud breathing can be distinguished by robust acoustic features. The findings might serve as a guidance to find a consensus for an objective definition of snoring compared to loud breathing.

I. INTRODUCTION

What is snoring? Almost everybody knows how snoring sounds and will recognise it when hearing it.

In the past decades, several attempts have been made to describe snoring by its acoustic properties [1], [2], [3]. In 1990, the American Sleep Disorders Association (ASDA), the predecessor organisation of today's American Academy of Sleep Medicine (AASM), defined snoring as "loud upper airways breathing . . . caused by vibrations of the pharyngeal tissues" [4]. In 1996, Dasmasso et al. noted that "snoring is a symptom of nasal obstruction . . . however, its acoustic features in these disorders are not well-defined" [5]. The authors defined a snoring index (numbers of snores per hour of sleep) and a snoring frequency (numbers of snores per minute of snoring time). Both definitions, however, refer to the frequency and severity of the snoring phenomenon, and do not consider the acoustic particularities of the snore sound itself.

More recently, in 2017, Swarnkar et al. described snoring as being "characterized by repetitive packets of energy that are responsible for creating the vibratory sound peculiar to snorers" [6]. This definition is based on the fact that, in most subjects, snoring is generated in the inspirational phases during successive, regular breathing cycles.

Nevertheless, no definition exists to date that permits an objectively measurable distinction between snoring and loud breathing, which can occur at very similar temporal patterns. As the Sleep Medicine Working Group of the German Society of Otorhinolaryngology, Head and Neck Surgery puts it: "To date, a satisfactory definition of snoring is lacking" [7]. Such a definition, however, is a fundamental prerequisite to develop algorithms that attempt to acoustically detect snore events during natural or artificial sleep. Several of such algorithms have been described [8], [9], [10], [11], [12]. Nevertheless, the distinction of snore and non-snore sounds has been made by the investigating authors themselves based on their own subjective judgement, making their findings not independently verifiable.

Rohrmeier et al. [13] made efforts to overcome this lack of an objective distinction between "snoring" and "loud breathing". In order to arrive at a reliable differentiation, they have created a corpus of nightly breathing and snore sounds which was classified by 25 human raters as either breathing or snoring. Although still based on subjective judgement, the high number of independent raters provides a certain common ground. The sounds were analysed for sound pressure level as well as for the psychoacoustic parameters loudness, sharpness, roughness, fluctuation, and annoyance. Annoyance yielded a sensitivity and specificity of 76.9 % and 78.8 %, respectively.

The aim of this work is to find more selective and more robust objective acoustic descriptors and to deploy machine learning methods for the distinction between snoring and breathing. These findings can later be used to develop and improve applications for automatic identification of snoring events during sleep.

The paper is structured as follows. Chapter II contains a description of the steps taken for preparation of the data and the machine learning methods used. The results are summarised in Chapter III. Discussion and a conclusion follow in Chapters IV and V.

[1]C. Janott and W. Hemmert are with the Munich School of BioEngineering, Technical University of Munich, Garching, Germany, c.janott@gmx.net.

[2]C. Rohrmeier is with the Faculty of Medicine, University of Regensburg, Regensburg, Germany.

[3]M. Schmitt is with the ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany.

[3,4]B. Schuller is with the Department of Computing, Imperial College London, London, UK, and the ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany.

Conflict of interest: The corresponding author holds a German patent on a method and system for the determination of anatomical causes of snoring noise (DE102012219128B4). All other authors declare no conflict of interest.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

II. MATERIALS AND METHODS

A. Database Properties and Data Preparation

The corpus created by Rohrmeier et al. comprises 55 audio sequences of nightly breathing and snore sounds from 23 subjects recorded during natural sleep in a sleep laboratory. The audio sequences are approximately 10 seconds in length, each sample containing three complete, consecutive respiratory cycles (inspiration and expiration). Care has been taken to include sounds that cover the whole spectrum from "normal" breathing to "heavy" snoring. The sounds were classified by 25 human raters. An inter-rater agreement of 75 % was used as a threshold to classify sounds as either "snoring" or "breathing". 16 percent of the sound sequences could not be classified unequivocally (inter-rater agreement of less than 75 %) and were labeled as "unclear". For details on subjects and annotation methods please refer to [13].

For our analysis, we have cut each of the 55 sequences into three separate segments only containing the inspiratory phase of the respective breathing cycle, as, in the predominant number of cases, snoring occurs during inspiration. Two exceptions have been made: the subjects in two of the recordings showed pronounced snoring during expiration, with an acoustically unobtrusive inspiration phase. In these cases, we selected the expiratory phase for analysis. Further, four samples were excluded as they contained a level of distortion that might negatively affect the extraction of acoustic features.

All segments have been normalised and stored in wav PCM format at 48 kHz sampling rate and 16 bit resolution. In total, the resulting database comprises 161 snore or breath samples with an average length of 1.88 s (range 0.53 ... 2.88 s). Of these, 95 samples were classified as "snoring" (S), 39 samples as "breathing" (B), and 27 samples as "unclear" (U).

The S-class and B-class samples were stratified into two sequence-disjunctive partitions, namely, a training and a development set together containing the samples from 45 sequences. All samples stemming from one sequence have always been assigned to the same partition. Because of the lack of an unequivocal label, the U-class samples were not included in either the training or the development set.

B. Machine Learning Experiments

For feature extraction, we used the OPENSMILE open-source audio feature extractor [14], [15]. We deployed the INTERSPEECH COMPARE feature set, based on 65 low level descriptors (LLDs), describing the temporal and spectral properties of the signal, as well as the first order derivatives (deltas) of these LLDs. The set is comprised of several statistical functionals derived from the LLDs and their deltas, resulting in a total number of 6 373 features. For a detailed description of the feature properties, please refer to [16] and [17]. The INTERSPEECH COMPARE feature set has been successfully used in a number of earlier experiments on the classification of snore sounds [18], [19], [20], [21].

The open-source support vector machine toolkit LIBLINEAR [22] was chosen to train a classifier. We compared the performance of two solver types: dual L2-regularised L2-loss support vector classification, and dual L2-regularised logistic regression. Linear SVMs achieve good results especially with smaller data sets and a large number of features, as is the case in our experiments. Furthermore, their generalisation behaviour can be well controlled by the complexity parameter, avoiding over-adaptation to the training data. OPENSMILE and the COMPARE feature set have yielded very good results in a number of earlier works [16], [17], hence we deployed it for these experiments.

TABLE I

FEATURE SUBSETS USED. *#LLDs*: NUMBER OF LOW-LEVEL DESCRIPTORS; *#Features*: NUMBER OF FEATURES WITH FIRST ORDER DERIVATIVES AND STATISTICAL FUNCTIONALS.

Subset	#LLDs	#Features	Description
audSpec	26	2600	Mel frequency spectrum-generated perceptual linear predictive cepstral coefficients
audspec	1	100	Sum of the audSpec coefficients
audspecRasta	1	100	Relative spectral transform-style filtered auditory spectrum
pcm_RMSenergy	1	100	Root mean square energy
pcm_zcr	1	100	Zero crossing rate
pcm_fftMag	15	1500	Magnitude of fast fourier transform coefficients
mfcc	14	1400	Mel-frequency cepstral coefficients
F0final	1	83	Smoothed fundamental frequency contour
voicingFinalUncl	1	78	Voicing probability
jitterLocal	1	78	Frame-to-frame period lengths differences between pitch periods
jitterDDP	1	78	First order derivative of jitter
shimmerLocal	1	78	Frame-to-frame amplitude differences between pitch periods
logHNR	1	78	Logarithmic ratio of harmonic signal energy to noise signal energy
COMPARE	130	6373	All subsets combined

In a first experiment, we performed a 45-fold cross validation using the S-class and B-class samples, each time leaving the samples of one sequence out of the training, respectively the development set, and used them for testing. The complexity parameter was set to 1, which has been experimentally determined as providing the optimal unweighted average recall (UAR) when optimised on the development partition in the range of $2^{-30}, 2^{-29}, \dots, 2^{0}$. Training of the final model was performed fusing the training and development partition, in each case without the samples of the respective testing sequence.

The experiments were carried out 14 times. Besides the full INTERSPEECH ComParE feature set, we deployed 13 subsets in order to determine those feature classes which are most sensitive for the distinction between snore and breath sounds. The subsets used and a description of the feature characteristics are listed in Table I.

TABLE II

CLASSIFICATION RESULTS PER FEATURE SUBSET OF THE S-CLASS AND B-CLASS SAMPLES USING TWO DIFFERENT SOLVER TYPES. *UAR*: UNWEIGHTED AVERAGE RECALL; *WAR*: WEIGHTED AVERAGE RECALL.

		SVM classification		Logistic regression	
No	Name of Set	% UAR	% WAR	% UAR	% WAR
1	audSpec	87.2	85.1	87.2	85.1
2	**audspec**	**92.2**	**93.3**	**91.7**	**92.5**
3	audspecRasta	62.3	67.9	60.2	67.2
4	pcm_RMSenergy	87.9	90.3	90.4	91.8
5	pcm_zcr	75.4	79.1	75.0	80.6
6	**pcm_fftMag**	**93.0**	**93.3**	**93.0**	**93.3**
7	mfcc	88.0	85.0	89.0	86.6
8	F0final	63.3	69.4	64.7	72.4
9	voicingFinalUncl	84.5	86.6	81.1	82.8
10	jitterLocal	67.4	73.1	64.9	71.6
11	jitterDDP	69.5	73.9	73.1	76.7
12	shimmerLocal	79.4	83.6	81.9	85.1
13	logHNR	81.9	85.1	79.6	82.8
14	**ComParE**	**92.9**	**91.0**	**92.4**	**90.3**

In a second experiment, we compared the probability values from the logistic regression solver type training results with the level of agreement of the human raters, i. e., the percentage of raters that defined the sounds of the respective sequence as snoring. An agreement of >75 % was defined as snoring (S), <25 % as breathing (B), between 25 % and 75 % as unclear (U). In this experiment, we used the data for the S-class and B-class type samples which were generated as described above. In addition, we used the full combined training and development partitions of all S-class and B-class samples for model training and tested the model on the U-class.

C. Ranking of Single Features

In order to evaluate the single most sensitive features for the distinction between snoring and breathing, we calculated for each of the 6 373 features the UAR, defined as the mean of the class-specific recalls for S-class and B-class samples. This exercise was done for all possible values of the respective feature, and considering the value yielding the maximum UAR as the ideal separator for this feature.

III. Results

The results of the first experiment are summarised in Table II. We used Unweighted Average Recall (UAR) as performance measure. The best classification performance could be achieved using the *pcm fftMag* feature subset, comprising 15 coefficients and their derivative and statistical functionals derived from the magnitudes (the real parts) of a fast Fourier transformation (FFT) of the signal. Both SVM classification and logistic regression yielded a UAR of 93.0 %. The second best performance showed the *audspec* feature subset, with a UAR of 92.2 % using SVM classification, and 91.7 % using logistic regression. Interestingly, this subset is based only on a single LLD, which is the sum of 26 perceptual linear predictive (PLP) cepstral coefficients generated from the mel frequency spectrum. The full INTERSPEECH ComParE feature set yielded a UAR of 92.9% using SVM classification, and 92.4% with logistic regression. Table III shows the confusion matrices using SVM classification for the three best-performing feature subsets.

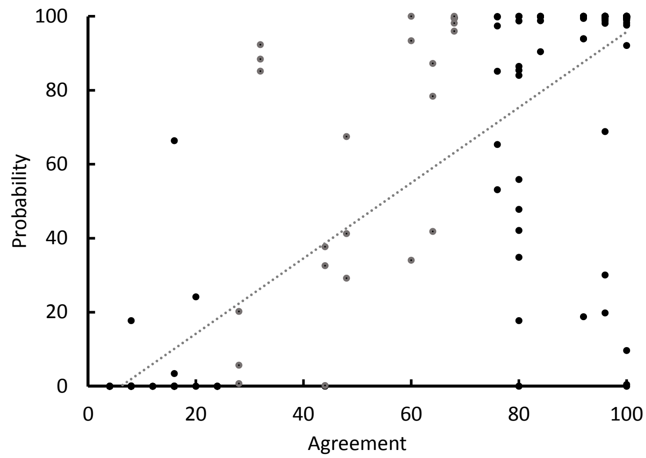

Fig. 1. Probabilities calculated by logistic regresssion versus inter-rater agreement. *x-axis:* Inter-rater-agreement (in %); *y-axis:* Probability values for the snoring class of the logistic regression model (in %). U-class samples are depicted in grey colour. The dashed line is the trendline.

Figure 1 shows a scatter plot of the probabilities from the trained logistic regresssion model versus the the percentage of raters that defined the sounds of the respective sequence as snoring (second experiment). Comparing the determination coefficient R^2 for all feature sets, we found that the full INTERSPEECH ComParE set yielded the best result with an R^2 of 0.66.

TABLE III

CONFUSION MATRICES OF THE BEST-PERFORMING FEATURE SUBSETS USING SVM CLASSIFICATION

audspec	pred ->	- S -	- B -
	- S -	97.4 %	5.3 %
	- B -	10.3 %	89.7 %

pcm_fftMag	pred ->	- S -	- B -
	- S -	93.7 %	6.3 %
	- B -	7.7 %	92.3 %

ComParE	pred ->	- S -	- B -
	- S -	88.4 %	11.6 %
	- B -	2.6 %	97.4 %

For the single features, Figures 2, 3, and 4 show scatter plots of inter-rater agreement versus value after openSMILE feature extraction of the three single features yielding the highest UAR. The x-axis shows the inter-rater agreement, the

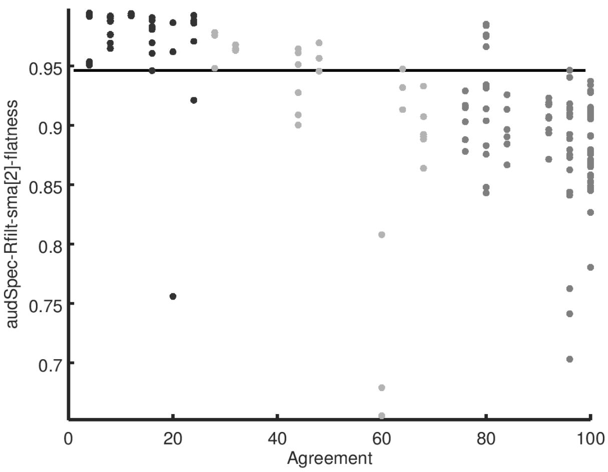

Fig. 2. Agreement vs feature *audSpec-Rfilt-sma[2]-flatness*.

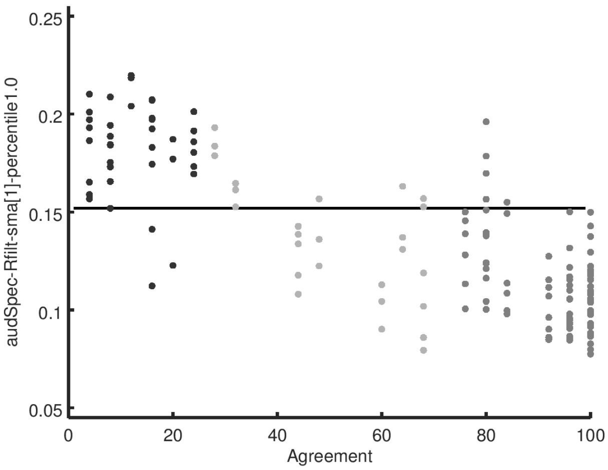

Fig. 3. Agreement vs feature *audSpec-Rfilt-sma[1]-percentile1.0*.

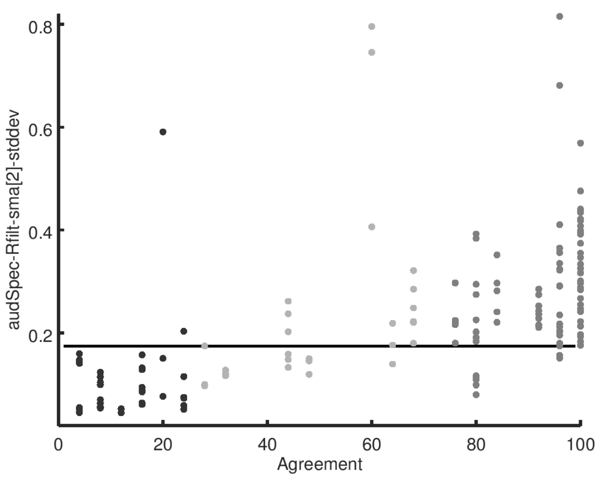

Fig. 4. Agreement vs feature *audSpec-Rfilt-sma[2]-stddev*.

y-axis displays the respective feature value. The horizontal line denotes the ideal separator (value of highest UAR).

Table IV summarises the UAR, Sensitivity and Specificity of the best-performing single features.

All of the three features are statistical functionals of a PLP cepstral coefficient generated from the mel frequency spectrum. Namely, the flatness of the second audspec coefficient (Figure 2), the 1%-percentile of the first coefficient (Figure 3), and the standard deviation of the first coefficient (Figure 4).

IV. Discussion

The best-performing single features as well as the second-best-performing feature subset are based on mel frequency spectrum-generated PLP cepstral coefficients. Perceptual linear prediction is very similar to linear predictive coding, with the difference that the PLP coefficient's spectral characteristics are matched to the characteristics of the human auditory system [23].

The best performing feature subset is based on FFT-generated features, whereas single FFT-based features yielded a UAR of up to 91.3 %, which is the next best performance after the audspec-based features.

AudSpec and FFT are different representations of the signal's spectral properties. By comparison, features describing the temporal properties, such as jitter and shimmer, did not prove to be as good predictors for the difference of snoring and breathing.

TABLE IV

Unweighted Average Recall (*UAR*), Sensitivity (*Sens*) and Specificity (*Spec*) of the best-performing single features.

Score	Name of Feature	% UAR	% Sens	% Spec
1	audSpec-Rfilt-sma[2] flatness	93.8	92.6	94.9
2	audSpec-Rfilt-sma[1] percentile1.0	93.5	94.7	92.3
3	audSpec-Rfilt-sma[2] stddev	93.2	91.6	94.9

Interestingly, based on our dataset, single features showed a performance that is comparable to models learnt on SVM-classification and logistic regression and based on complete feature sets. The generalisability of classifications based on a single feature remains questionable, however, and might well be worse than a feature set-based machine learnt model when applied to unknown, independent data.

Our results are notably better than those reported by Rohrmeier et al. using psychoacoustic parameters. Rohrmeier et al. found that annoyance according to Zwicker's psychoacoustic annoyance model yielded the best distinction between loud breathing and snoring (sensitivity 76.9 %, specificity 78.8 %). Zwicker's annoyance model combines four acoustical parameters, i. e., loudness, sharpness, fluctuation strength and roughness [24]. Loudness is derived from the sound pressure level (SPL), which in turn depends on the distance and position of the recording microphone relative to the sound source, i. e., the snorer's mouth and nose. This parameter therefore requires a careful setup and calibration of the recording situation. Differences in microphone positions, amplification settings of the recording equipment, and even sleeping positions of the snoring subject may result in differences of the annoyance value and therefore skew the results.

The PLP cepstral coefficients, in contrast, are based on the spectral properties of the signal and independent of the absolute SPL. Further, the amplitude of the audio samples has been normalised. This promises to yield more robust results when used in real life applications, where microphone positions and room conditions might not be precisely controllable.

A weaknesses of this study is that our experiments are based on a ground truth that is still subjective, although the high number of raters promises a certain level of consensus compared to classifications that are based on the evaluation of one single or a small group of raters. Further, the original classification by the raters was made by listening to all three snore cycles of the respective sequence. For our experiments, we separated these into single samples. Potential differences in sound between the three respiratory cycles of the same individual have therefore not been considered, a fact that potentially might have skewed our results. Finally, the size of the corpus is small for a machine learning task. Therefore, the robustness and generalisability of our findings is yet to be proven by larger datasets.

V. Conclusion and Outlook

We could show for the first time that snoring and loud breathing can be distinguished by acoustic features that represent the spectral properties of the signal and that are independent of the SPL. This work might help to find a consensus for an objective definition of snoring as opposed to (loud) breathing based on acoustic parameters, and serve as a guidance for future applications in the automatic detection of snore sounds. Subject of future work will focus on increasing both the dataset size and the number of independent raters, as well as adding other typical breath-related sounds, such as wheezing or moaning. This way, an even more differentiated distinction of sounds can contribute to automated acoustic analysis of human sleep quality and sleep disorders.

References

[1] J. Schafer, "How can one recognize a velum snorer?" *Laryngorhinootologie*, vol. 68, no. 5, pp. 290–294, May 1989.

[2] D. Pevernagie, R. M. Aarts, and M. De Meyer, "The acoustics of snoring," *Sleep Medicine Reviews*, vol. 14, no. 2, pp. 131–144, 2010.

[3] H. Alshaer, A. Pandya, T. D. Bradley, and F. Rudzicz, "Subject independent identification of breath sounds components using multiple classifiers," in *Proceedings of the IEEE ICASSP 2014*. Florence, Italy: IEEE, 2014, pp. 3577–3581.

[4] M. Thorpy, *The international classification of sleep disorders: diagnostic and coding manual*. Lawrence KS ed Allen Press, USA, 1990.

[5] F. Dalmasso and R. Prota, "Snoring: analysis, measurement, clinical implications and applications," *European Respiratory Journal*, vol. 9, no. 1, pp. 146–159, 1996.

[6] V. Swarnkar, U. Abeyratne, and R. Sharan, "Automatic picking of snore events from overnight breath sound recordings," in *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Seogwipo, South Korea: IEEE, July 2017, pp. 2822–2825.

[7] B. Stuck, A. Dreher, C. Heiser, M. Herzog, T. Kühnel, J. Maurer, H. Pistner, H. Sitter, A. Steffen, and T. Verse, "Sk2 guidelines diagnosis and therapy of snoring in adults compiled by the sleep medicine working group of the german society of otorhinolaryngology, head and neck surgery," *HNO*, vol. 61, no. 11, pp. 944–957, 2013.

[8] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, Y. Serinagaoglu, and T. Akcam, "An efficient method for snore/nonsnore classification of sleep sounds," *Physiological Measurement*, vol. 28, no. 8, pp. 841–853, 2007.

[9] K. Qian, Z. Xu, H. Xu, and B. P. Ng, "Automatic detection of inspiration related snoring signals from original audio recording," in *Proceedings of the IEEE ChinaSIP 2014*. IEEE, 2014, pp. 95–99.

[10] K. Qian, Z. Xu, H. Xu, Y. Wu, and Z. Zhao, "Automatic detection, segmentation and classification of snore related signals from overnight audio recording," *IET Signal Processing*, vol. 9, no. 1, pp. 21–29, 2015.

[11] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins, "Silence-breathing-snore classification from snore-related sounds," *Physiological Measurement*, vol. 29, no. 2, pp. 227–243, 2008.

[12] A. Yadollahi and Z. Moussavi, "Automatic breath and snore sounds classification from tracheal and ambient sounds recordings," *Medical Engineering & Physics*, vol. 32, no. 9, pp. 985–990, 2010.

[13] C. Rohrmeier, M. Herzog, T. Ettl, and T. Kuehnel, "Distinguishing snoring sounds from breath sounds: a straightforward matter?" *Sleep and Breathing*, vol. 18, no. 1, pp. 169–176, 2014.

[14] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM International Conference on Multimedia*. Florence, Italy: ACM, 2010, pp. 1459–1462.

[15] F. Eyben, F. Weninger, F. Groß, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM International Conference on Multimedia*. Barcelona, Spain: ACM, 2013, pp. 835–838.

[16] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism," in *Proceedings of INTERSPEECH*, Lyon, France, 2013, pp. 148–152.

[17] F. Weninger, F. Eyben, B. Schuller, M. Mortillaro, and K. Scherer, "On the acoustics of emotion in audio: What speech, music and sound have in common," *Frontiers in Emotion Science*, vol. 4, pp. 1–12, 2013.

[18] K. Qian, C. Janott, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "Classification of the Excitation Location of Snore Sounds in the Upper Airway by Acoustic Multi-Feature Analysis," *IEEE Trans Biomed Eng*, Oct 2016.

[19] B. Schuller, S. Steidl, A. Batliner, E. Bergelson, J. Krajewski, C. Janott, A. Amatuni, M. Casillas, A. Seidl, M. Soderstrom, A. S. Warlaumont, G. Hidalgo, S. Schnieder, C. Heiser, W. Hohenhorst, M. Herzog, M. Schmitt, K. Qian, Y. Zhang, G. Trigeorgis, P. Tzirakis, and S. Zafeiriou, "The interspeech 2017 computational paralinguistics challenge: Addressee, cold & snoring," in *Proceedings of INTERSPEECH*, Stockholm, Sweden, 2017, pp. 20–24.

[20] C. Janott, M. Schmitt, Y. Zhang, K. Qian, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "Snoring classified: The munich-passau snore sound corpus," *Computers in Biology and Medicine*, vol. 94, no. March, pp. 106–118, 2018.

[21] K. Qian, C. Janott, J. Deng, C. Heiser, W. Hohenhorst, M. Herzog, N. Cummins, and B. Schuller, "Snore sound recognition: On wavelets and classifiers from deep nets to kernels," in *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, South Korea, 2017.

[22] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin, "Liblinear: A library for large linear classification," *Journal of machine learning research*, vol. 9, no. August, pp. 1871–1874, 2008.

[23] H. Hermansky, "Perceptual linear predictive (plp) analysis of speech," *The Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738–1752, 1990.

[24] H. Fastl and E. Zwicker, *Psychoacoustics, Facts and Models*. Springer-Verlag, 2007.