This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

Classification of the Excitation Location of Snore Sounds in the Upper Airway by Acoustic Multi-Feature Analysis

Kun Qian, *Student Member, IEEE,* Christoph Janott, *Student Member, IEEE,* Vedhas Pandit, *Student Member, IEEE,* Zixing Zhang, *Member, IEEE,* Clemens Heiser, Winfried Hohenhorst, Michael Herzog, Werner Hemmert, *Senior Member, IEEE* and Björn Schuller, *Senior Member, IEEE*

Abstract*—*Objective: **Obstructive Sleep Apnea (OSA) is a serious chronic disease and a risk factor for cardiovascular diseases. Snoring is a typical symptom of OSA patients. Knowledge of the origin of obstruction and vibration within the upper airways is essential for a targeted surgical approach. Aim of this paper is to systematically compare different acoustic features, and classifiers for their performance in the classification of the excitation location of snore sounds.** ***Methods:*** **Snore sounds from 40 male patients have been recorded during Drug-Induced Sleep Endoscopy, and categorized by ENT experts. Crest Factor, Fundamental Frequency, Spectral Frequency Features, Subband Energy Ratio, Mel-Scale Frequency Cepstral Coefficients, Empirical Mode Decomposition-Based Features, and Wavelet Energy Features have been extracted and fed into several classifiers. Using the ReliefF algorithm, features have been ranked and the selected feature subsets have been tested with the same classifiers.** ***Results:*** **A fusion of all features after a ReliefF feature selection step in combination with a Random Forests classifier showed the best classification results of 78 % Unweighted Average Recall by subject independent validation.** ***Conclusion:*** **Multi-feature analysis is a promising means to help identify the anatomical mechanisms of snore sound generation in individual subjects.** ***Significance:*** **This paper describes a novel approach for the machine-based multi-feature classification of the excitation location of snore sounds in the upper airway.**

***Index Terms*—Obstructive Sleep Apnea, Snore Sound Classification, Multi-Feature Analysis, Drug-Induced Sleep Endoscopy.**

K. Qian is with the Machine Intelligence & Signal Processing group, MMK, Technische Universität München, Arcisstr. 21, 80333, Munich, Germany (e-mail: andykun.qian@tum.de).

C. Janott and W. Hemmert are with the Institute for Medical Engineering, Technische Universität München, Arcisstr. 21, 80333, Munich, Germany (e-mail: c.janott@gmx.net; werner.hemmert@tum.de).

V. Pandit and Z. Zhang are with the Chair of Complex & Intelligent Systems, University of Passau, Innstr. 43, 94032, Passau, Germany (e-mail: vedhas@gmail.com; zixing.zhang@uni-passau.de).

C. Heiser is with the Department of Otorhinolaryngology/Head and Neck Surgery, Technische Universität München, Ismaningerstr. 22, 81675, Munich, Germany (e-mail: clemens.heiser@tum.de).

W. Hohenhorst is with the Clinic for ENT Medicine, Head and Neck Surgery, Alfried Krupp Krankenhaus, Alfried-Krupp-Str. 21, 45131, Essen, Germany (e-mail: mail@hohenhorst.com).

M. Herzog is with the Clinic for ENT Medicine, Head and Neck Surgery, Carl-Thiem-Klinikum Cottbus, Thiemstr. 111, 03048 Cottbus, Germany (e-mail: M.Herzog@ctk.de).

B. Schuller is with the Department of Computing, Imperial College London, 180 Queens' Gate, Huxley Bldg., London SW7 2AZ, UK, and the Chair of Complex & Intelligent Systems, University of Passau, Innstr. 43, 94032, Passau, Germany (e-mail: schuller@ieee.org).

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

I. Introduction

WITH a prevalence of 13 % (men) and 6 % (women) in the US population [1], Obstructive Sleep Apnea (OSA) is a chronic disease that can severely affect health and quality of life. OSA is defined as a syndrome with cessation or reduction of airflow during sleep due to complete (apnoea) or partial (hypopnea) collapse of the upper airway for more than ten seconds and with five or more episodes per hour in sleep [2]. It is usually associated with a decrease in oxy-hemoglobin saturation [2]. When untreated, OSA can among other symptoms, result in daytime sleepiness and morning headache [3]. Furthermore, it is an independent risk factor for cardiovascular diseases, stroke, hypertension, myocardial infarction, and is associated with diabetes and vulnerability to accidents [4], [5].

Loud snoring, as a typical symptom of OSA, is reported in more than 80 % of OSA patients [6]. The acoustic properties of snoring have been analyzed by researchers in acoustics and otorhinolaryngology with the aim of developing methods to replace or complement the gold standard for the diagnosis of OSA, Polysomnography (PSG) [7]. Works by pioneers have shown that methods based on snoring sound analysis can reach the sensitivities and specificities up to 90 % and accuracies up to 80 % in the detection of OSA [8]. Even though based on small populations (generally between 5 and 60 subjects) [8], the results are promising and encouraging.

Due to the multifactorial mechanisms of snore sound (SnS) generation, and depending on the individual anatomy, surgical options for OSA differ and include, among others, tonsillectomy or tonsillotomy, uvulotomy, uvulopalatopharyngoplasty, soft palate stiffening, tongue base suspension, hypoglossal nerve stimulation, mandibular advancement, epiglottectomy, and hyoid suspension [9]. Especially in severe OSA, a combination of several surgical treatments at different anatomic levels (multilevel surgery) [10] is often used. The analysis of the individual anatomical site of snoring sound generation, and of the obstruction mechanism can lead to a targeted, and less invasive surgical approach.

Drug induced sleep endoscopy (DISE) is increasingly used to identify the location and form of vibrations and obstructions [11]. However, DISE is time consuming, costly, and straining for patients. Further, it cannot be performed in natural sleep. Another method to identify the location of obstruction in

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

the upper airway is multi-channel pressure measurement [12], [13], [14], [15]. Here, a thin tube with multiple pressure sensors is introduced into the upper airway. The pattern of pressure changes during breathing of the different sensors allows a determination of the obstruction location during an apnoeic or hypopnoeic event. An advantage of this method is that, it can be used in natural sleep. However, the tube within the upper airway is not tolerated by every patient. Acoustic analysis could be an alternative to determine the vibration mechanisms within the upper airway, which is easier for doctors and patients.

Fewer studies exist on how to determine the location, and form of vibration and obstruction in the upper airway from the acoustic properties of snore sounds. Miyazaki et al. [16] adopted fundamental frequency to distinguish SnS generated by soft palate, tonsils/tongue base, combined type (both palate, and tonsils/tongue base), and the larynx. Based on the examination of 75 adult patients they concluded that, the average value of fundamental frequency was 102.8 Hz, 331.7 Hz, 115.7 Hz and around 250 Hz in the corresponding sites mentioned above, respectively. Hill et al. found the crest factor, the ratio of peak to root mean square value of a time-varying signal, to be significantly higher for palatal snorers ($p < 0.01$, Student-t or Mann-Whitney tests) in 11 patients [17]. Agrawal et al. found that, the SnS generated by palate and tongue were respectively characterized by low and high peak frequency (137 Hz and 1 243 Hz), while epiglottic snores occured at 490 Hz and tonsillar snores at 170 Hz within a population of 16 subjects [18]. Beeton et al. proposed a combination of a 2-means clustering method and the statistical dimensionless moment coefficients of skewness and kurtosis to discriminate palatal, and non-palatal SnS collected from 15 patients [19]. They indicated that, the statistical moment coefficients demonstrate a method of measurement of the peakedness and symmetry of the impulse.

The studies mentioned above are focused on evaluating certain well-selected acoustic features for their sensitivity to the anatomical mechanisms of snoring sound generation or upper airway obstruction. The comparisons, and results are based on statistical analysis, and basic signal observation. The application of multi-feature analysis methods to this problem has been proposed [20]; however, advanced signal processing methods, and machine learning models have not yet been used for this purpose. In our approach, we combine the methods used previously, and introduce new features within advanced classification techniques. The structure of this paper is as follows: In Section II we present the data acquisition system and the methods for multi-feature analysis-based classification. Detailed experimental results are given in Section III. Finally, we discuss the findings in Section IV, and provide a conclusion in Section V.

II. Materials and Methods

A. Snore Sounds Acquisition and Labeling

This study is approved by the ethic committee of Klinikum rechts der Isar, Technische Universität München, Germany. We used SnS data from 40 male subjects which were diagnosed with primary snoring or OSA through a Polysomnography (PSG). The demographic data for the subjects is shown in Table I.

TABLE I
Demographic information of subjects. BMI: Body Mass Index; AHI: Apnea Hypopnea Index.

	MEAN	STD. DEV.	RANGE
Age (*years*)	47.4	11.5	26 – 71
BMI (kg/m^2)	26.9	3.1	21.2 – 38.4
AHI (*events/h*)	21.7	12.8	1.3 – 59.1

In addition to PSG, DISE was performed in all subjects in order to determine adequate surgical intervention measures. DISE videos were recorded at Klinikum rechts der Isar, Munich, Germany, at Alfried Krupp Hospital, Essen, Germany, and at University Hospital Halle (Saale), Germany, using a flexible nasopharyngoscope (see Fig. 1 as an example of the clinical setting). Audio information was recorded in parallel using a headset microphone (in Munich), or a handheld mircrophone (in Essen, and Halle), respectively, and synchronously stored in the same file. Based on the video and audio recordings, the locations of sound generation were categorized by an ENT expert based on the VOTE classification [21]. VOTE is a popular classification which distinguishes four levels within the upper airway: the level of the velum (V), the oropharyngeal area including the palatine tonsils (O), the tongue base (T), and the epiglottis (E) (see Fig. 2). Only recordings that showed a clearly identifiable, single source of snoring sound have been included. Snoring events with mixed forms (several vibration locations) or unclear source of vibration were excluded. From each included recording, three to five snoring events, which showed no obstructive disposition, have been manually selected. These snoring events have then been extracted from the audio data stream, and labelled based on the VOTE classification. In fact, 'snore site' and 'obstruction site' in the upper airway are two different definitions, which may or may not coincide in individual patients. In this study, we exclusively focus on the determination of the site of vibration as a cause for the generation of snore sounds.

Of the 40 subjects, 11, 11, 8, and 10 subjects were categorized to be V, O, T, and E-type snorers, respectively . Between one and five snoring *events* per type were extracted per subject. In total, we used 164 snoring events (41 episodes for each type of SnS, length ranging from 0.728 to 2.495 s with an average of 1.498 s). We segmented the events into single *segments* for further feature extraction and machine learning. Every segment has a length of 200 ms and neighbouring segments have an overlap of 50 %. We performed a subject-independent validation to evaluate the performance of our trained classifiers. As indicated by Roebuck et al. [8], previous works on snoring audio analysis have not been based on independent test data sets. In order to achieve substantiated results with a practical relevance, we use subject-independent testing sets in our study. We randomly separated the 40 subjects' data into the train, development (dev), and test sets within the proportion of 60 %, 20 %, and 20 % of the total data set. The number of segments and independent subjects for each set are shown in Table II.

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

Fig. 1. Example of the DISE clinical setting in Munich. The video of the upper airway was recorded using a flexible nasopharyngoscope (Storz, Germany at the Munich and Halle sites and Olympus, Germany, at the Essen site) connected to a video recording system (Telepack X, Storz, Germany, at the Munich site; AIDA, Storz, Germany, at the Halle site; rpSzene, Rehder/Partner, Hamburg, Germany, at the Essen site). The audio signal was simultaneously recorded using a microphone connected to the same recording system. Audio and video information was stored in the same file.

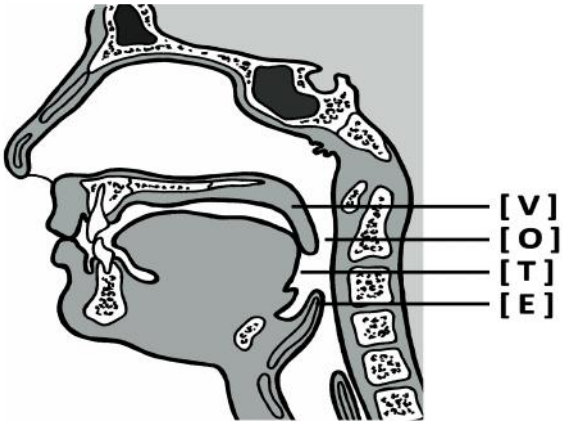


Fig. 2. Corresponding positions of the VOTE classification in the upper airway. 'V' represents the level of the velum. 'O' represents the oropharyngeal area. 'T' represents the tongue base. 'E' represents the level of the epiglottis.

The segments are divided into *frames* of 64 ms length and an overlap of 50 %. Features and statistical functionals are applied to each frame in every segment and all attribute information is stored for the further machine learning steps.

TABLE II
NUMBER OF SEGMENTS (INDEPENDENT SUBJECTS) FOR EACH SNORE-TYPE AND DATA SET.

	train	dev	test	$\sum$
V-Type	363 (7)	104 (2)	152 (2)	619 (11)
O-Type	326 (7)	125 (2)	122 (2)	573 (11)
T-Type	289 (4)	90 (2)	78 (2)	457 (8)
E-Type	323 (6)	96 (2)	148 (2)	567 (10)
$\sum$	1 301 (24)	415 (8)	500 (8)	2 216 (40)

B. Feature Extraction

Until most recent, the lion's share of the work done in multi-feature snoring analysis is focused on finding and evaluating suitable acoustic feature sets [7] to identify obstructive events, to distinguish between primary snoring and OSA, or to estimate the severity of OSA. Abeyratne et al. proposed a multi-feature analysis method built on combined models of pitch, total airway response (TAR) estimators and Mel-frequency cepstral coefficients (MFCC, 0–12), and achieved 89.3 % sensitivity with 92.3 % specificity in OSA detection [22].

Fig. 3 illustrates the waveforms, and the corresponding spectrograms of typical V, O, T, and E type SnS episodes. We can see that, the main energy components in three of the classes are concentrated in the frequency area below around 5 000 Hz. Energy and spectral distribution characteristics are similar, except for the Type T, which shows higher energy content above 2 500 Hz compared to the other three.

Motivated by the results of the multi-feature analysis methodology [22], combined with the basic spectral analysis of the four types of SnS, we propose nine basic acoustic feature sets, and systematically explore, and compare their performance in the classification of snore sounds based on the VOTE system.

1) Crest Factor: Hill et al. in 1999 proposed the use of the *crest factor* [17], we re-define it as shown in our previous study [23]:

$$Crest\ Factor = \frac{V_{90}}{V_{rms}}, \tag{1}$$

where V_{90} is the 90^{th} centile maximum absolute value in the digitized sound epoch, and V_{rms} is the root mean square of the amplitude values (between the 10^{th} and 90^{th} centile maximum absolute values) in one epoch. Elimination of the lowest (below 10^{th}) and highest (above 90^{th}) values is done to minimize the impacts of both random and quantization noise.

2) F0: We estimate the fundamental frequency (F0) of SnS with an algorithm based on spectrum shifting on a logarithmic frequency scale and calculating the Subharmonic-to-Harmonic Ratio (SHR), which was proposed by Sun [24].

3) Formants: In our study, an 18th-order linear predictive coding (LPC) is performed to estimate the formants. Like in speech analysis, the LPC parameters are determined via the Yule-Walker autoregressive method along with the Levinson-Durbin recursive procedure [25]. Then the formant frequencies can be estimated from the angles of the positive values of the complex roots in an all-pole model as follows:

$$H_{(z)} = \frac{1}{1 - \sum_{q=1}^{p} \alpha_q z^{-q}}, \tag{2}$$

where α_q $(q = 1, 2, 3, ..., p)$ are the LPC parameters. Subsequently, we extract the first three formant frequencies (i. e., F1, F2, and F3), and the corresponding amplitude energies to create a formants feature set.

4) Spectral Frequency Features: The spectrum of SnS carries vital information on the state of the upper airway [26]. Certain frequency features such as peak frequency, center frequency, and mean frequency, were studied both on the diagnosis of OSA [22] and distinction of the snore site [18]. In our previous work [27], we found that, spectral frequency features (SFF) can achieve a good performance on the classification of inspiration related SnS. Here we define F_{center},

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

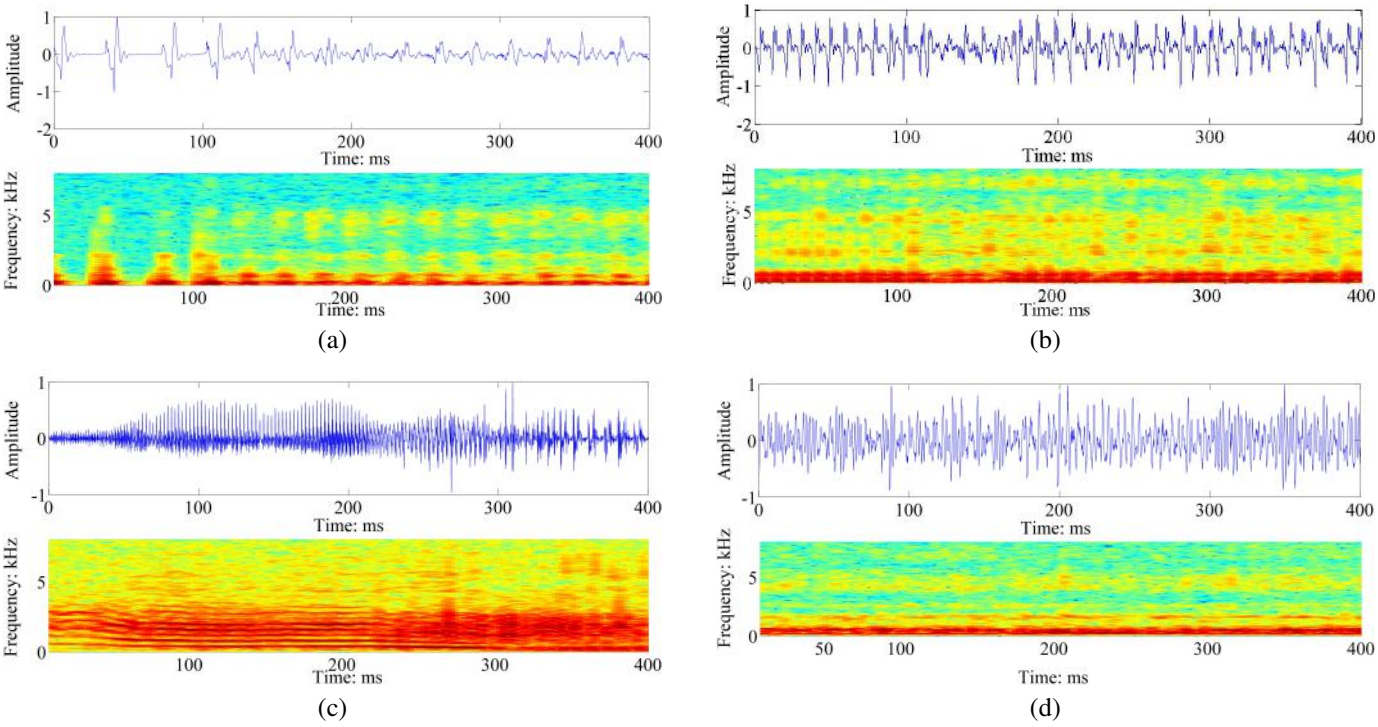


Fig. 3. Waveforms and spectrograms of typical VOTE SnS events. (a) V (velum) typical snoring event; (b) O (oropharyngeal) typical snoring event; (c) T (tongue base) typical snoring event; (d) E (epiglottis) typical snoring event.

and F_{peak} respectively as the half and maximum point in the full spectrum of the SnS in [18], F_{mean} is defined as:

$$F_{mean} = \frac{\sum_{f_i=0}^{f_c} f_i S(f_i)}{\sum_{f_i=0}^{f_c} S(f_i)}, \tag{3}$$

where f_c is the cut-off frequency of the SnS spectrum (in our study f_c is 8 kHz). $S(f_i)$ is the absolute amplitude of the spectrum at the frequency of f_i Hz. The spectrum calculations in this study are based on the Fast Fourier Transform (FFT). In addition, we define the F_{mean} of the 1 000 Hz sub-band spectrums as:

$$F_{mean(j)} = \frac{\sum_{f_i=1000(j-1)}^{1000j} f_i S(f_i)}{\sum_{f_i=1000(j-1)}^{1000j} S(f_i)}, \tag{4}$$

where $j = 1, 2, 3, ..., 8$. Thus, we obtain detailed information on the spectral energy distribution in the sub-bands of SnS.

5) Power Ratio: Power ratio compares the relative amount of power emanating below and above a set frequency [18]. Some researchers chose the frequency at 800 Hz [28], and others chose it to be 750 Hz [18]. In this study, we use the power ratio at the frequency of 800 Hz and define it as:

$$PR_{800} = \lg \frac{\sum_{f_i=0}^{800} (S(f_i))^2}{\sum_{f_i=800}^{f_c} (S(f_i))^2}, \tag{5}$$

6) Subband Energy Ratio: The subband energy ratio (SER) describes the relative energy distribution in subbands of the SnS spectrum. It had been demonstated to be efficient in snore/nonsnore classification [29], [30]. We extract a 1 000-Hz SER feature set as:

$$SER_{1000}(j) = \frac{\sum_{f_i=1000j-1}^{1000j} (S(f_i))^2}{\sum_{f_i=0}^{f_c} (S(f_i))^2}, \tag{6}$$

where $j = 1, 2, 3, ..., 8$.

7) Mel-Scale Frequency Cepstral Coefficients: Mel-scale frequency cepstral coefficients (MFCCs) have long been demonstated to be efficient features for speech recognition [31]. In our previous study [32], we found that, MFCCs can outperform other spectral features on classification of SnS. In this study, we extract thirteen Mel cepstral coefficients (MFCCs 0–12) obtained from SnS passing 27 triangular Mel filter banks.

8) Empirical Mode Decomposition-Based Features: SnS are a typical kind of non-stationary signal [7]. Therefore, FFT-based methods are not suitable to reveal detailed information on the variation of the SnS in the time domain. Empirical mode decomposition (EMD), based on choosing basis functions, is adaptive to characterize non-stationary signals [33]. Motivated by the performance of EMD-based features (EMDF) for classification of roller bearing fault vibration signals [34], we extract the subband EMD energy ratio $EMD_{ratio}(k) = E_k/E$. E_k is the energy (sum of squares) of the k-th level intrinsic mode functions (IMFs) decomposed by EMD from the SnS. E is the total energy of the whole SnS within EMD (the residual is eliminated). In addition, we calculate the entropy of the EMD_{ratio} as:

$$H_{EMD_{ratio}} = -\sum_{k=1} EMD_{ratio}(k) \lg(EMD_{ratio}(k)). \tag{7}$$

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

9) Wavelet Energy Features: It is known that, wavelet transform (WT) is a useful tool to analyze non-stationary signals [35]. In 2007, Matsiki et al. studied how to use WT-based methods to analyze SnS of OSA patients [36]. Ng et al. used wavelet transform to enhance the snore signal from a noisy environment for improving the feature extraction [37]. Khushaba et al. [38] presented a wavelet-packet-based feature extraction algorithm and adopted it to classify five different drowsiness levels based on EEG, EOG and ECG signals. In this study, we introduce this WPT method into SnS multi-feature extraction.

The core technique of the algorithm proposed by Khushaba et al. is the wavelet packet transform (WPT), introduced by Coifman et al. [39]. The WPT could be understood as a tree of subspaces, where $\Phi_{0,0}$ is the root node. The signal space $\Phi_{l,m}$ (l is the level of the decomposition process and m is the subband index) is decomposed into two orthogonal subspaces level by level: $\Phi_{l+1,2m}$ and $\Phi_{l+1,2m+1}$, namely, the approximation space, and the detail space [40].

This decomposition process is done by dividing an orthogonal basis $\Omega_l(t-2^l m)_{l,m\in Z}$ from $\Phi_{l,m}$ into two new orthogonal bases: $\Omega_{l+1}(t-2^{l+1}m)_{l,m\in Z}$ from $\Phi_{l+1,2m}$, and $\Psi_{l+1}(t-2^{l+1}m)_{l,m\in Z}$ from $\Phi_{l+1,2m+1}$, respectively, where, $\Omega_{l,m}(t)$, and $\Psi_{l,m}(t)$ are wavelet functions [35] (we select 'sym3'[1] of the 'Symlets' wavelet function family due to its best performance in our previous experiments [41]). A construction of normalized filter bank energy is defined as:

$$E_{\Phi_{l,m}} = \sqrt{\frac{\sum_n (\mathbf{w}_{l,m,n})^2}{N_m}},\ n\ =\ 0,1,2,...,2^{l-1}, \quad (8)$$

where $\mathbf{w}_{l,m}$ represent the WPT coefficients evaluated from the signal at the subspace $V_{l,m}$ and N_m is the number of wavelet coefficients in the m-th subband. Therefore, $E_{\Phi_{l,m}}$ denotes the normalized bank filter energy in m-th subband at the l-th decomposition level. Due to the origination of these features from WPT-based coefficients, we call them WPT Energy (WPTE) features.

As a complementary feature set, we use the wavelet transform (WT) to define a percentage-like WT-based Energy (WTE) feature as:

$$\hat{E}_{\hat{\Phi}_l} = \frac{(\hat{\mathbf{w}}_l)^2}{\sum_{l=1}^{L_{max}} (\hat{\mathbf{w}}_l)^2}\ \times\ 100, \quad (9)$$

where $\hat{\mathbf{w}}_l$ are the coefficients generated by WT at the l-th decomposition level. In addition, the variance, waveform length (the sum of absolute differences), and the entropy are calculated from the base by Eq. 9.

WT decomposes only the approximation part, using low pass filters (LPF), whereas WPT decomposes both the approximation, and the detail part (low pass and high pass filtering (HPF)). All WPTE and WTE features are modified with a logarithmic operator. We generate $2^{L_{max}+1}-1$ WPTE related features, and $4\times(L_{max}+1)$ WTE related features, where L_{max} is the maximum level for wavelet decomposition. We then fuse all WPTE and WTE features in one combined feature set. Since these are all energy related features, we call the resulting feature set 'Wavelet Energy Features (WEF)'. Here L_{max} is set to be 7, therefore, we extract $255+32=287$ descriptors in total.

[1]The wavelet function names in this paper are all identical to the ones given in the Wavelet Toolbox of Matlab by MathWorks® (http://www.mathworks.com/products/wavelet/).

TABLE III
FRAME-BASED FEATURE SETS AND STATISTICAL FUNCTIONALS.

Frame-based feature sets (338)	**Statistical functionals** (9)
Crest Factor (1), *F0* (1), *Formants* (6), *SFF* (11), PR_{800} (1), *SER* (8), *MFCCs* (13), *EMDF* (10), *WEF* (287)	max, min, mean, range, standard deviation, slope, bias (linear regression approximation) skewness, kurtosis

10) Statistical Functionals: In order to evaluate the non-stationary characteristics of the material, the differences between the frames within a segment are taken into account [7]. Motivated by the success of our large scale feature extraction toolkit, openSMILE [42], we implement statistical functionals into our SnS feature extraction. After calculating the frame-based features using the algorithms described above, the statistical functionals are applied to each frame in every segment. The whole attribute information of this segment is used for the machine learning process. Detailed information on this technique can be found in [42]. The basic frame-based feature sets and the statistical functionals are listed in Table III.

C. Machine Learning Models

SnS recognition based on the VOTE classification is a four-class classification task. We select and compare seven machine learning models which we have chosen based on their popularity, diversity, and abilities. *K*-Nearest Neighbors (*K*-NN), and Linear Discriminant Analysis (LDA) are chosen since their probabilistic models are built for and they are frequently applied in biomedical signal processing tasks [43]. Models with a mature theoretical foundation like Feedforward Neural Networks (FNN) [44]), Support Vector Machines (SVM)[2] [45], and Random Forests (RF) [46] (an ensemble classifier [47]) are also used in our experiments. Further, Extreme Learning Machines (ELM) [48], and the Kernel-ELM (KELM) [49], one recent popular fast, and accurate classifier models, are explored.

D. ReliefF Feature Selection

To achieve a higher classification accuracy and to understand how the different features contribute to the machine learning models, we add a feature selection phase into our classifier optimization process. Motivated by the success of our previous study [27], [32], we employ the ReliefF algorithm for feature ranking and selection. Unlike principle component analysis (PCA), ReliefF retains the original physical meaning of each feature set in the vector. That is, we can further use these information of the ranked features to find the relationship

[2]The SVM classifier is implemented by the frequently-used, and mature toolkit LIBSVM [50].

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

between feature properties and anatomical characters, which is significant in our study.

Proposed by Kira and Rendell [51], Relief is a feature selection algorithm used in binary classification, which is repeated t times to update the weight vector (initially it is set to be zeros) as follows:

$$W_a = W_a - (x_a - nearHit_a)^2 + (x_a - nearMiss_a)^2, \quad (10)$$

where $nearHit_a$ is the closest same-class segment within the a-th feature to a randomly selected segment x. Likewise, $nearMiss_a$ is the closest different-class segment. After t times, each element of W_a will be divided by t. Thus, the weight of any given feature will decrease if the distance of that feature to nearby segments of the same-class is longer than to nearby segments of a different-class, and increase in the reverse case. W_a can be regarded as a reward and punishment factor due to the classification performance of each feature: Features with a high W_a show a good performance for the classification task at hand. ReliefF is an extension of Relief which can handle multiple classes and performs better with noisy data. It searches for b (in our study, b is empirically set to be 5) nearest hits and misses and averages their contributions for updating W_a, weighted with the prior probability of each class [52]. Since W_a evaluates the quality of each feature, we can rank the features according to their performance, and select the best ones to construct a new subset of the original features. We define the *Rank Ratio* as:

$$Rank\ Ratio = \frac{\sum_{a=1}^{M} W_a^+}{\sum W_a^+}, \quad (11)$$

where W_a^+ represents the positive weights of the features sorted in a descending order and M is the number of features included in the subset. The features with negative weights (W_a^-) are eliminated. By testing the classifier with feature subsets of different sizes (based on different *Rank Ratio*s), a feature subset for optimal classification performance can be identified, while at the same time, the size of the required feature set can be reduced.

III. Experiments and Results

A. Experimental Setup

All experiments are done within the software environment of Matlab R2016a by MathWorks®. The grids of main parameters for each classifier model mentioned in Section II-C are listed in Table IV. We chose these parameters by empirical experiments to optimize the performance of classification on the development data. A statistical significance p value is calculated by a one-sided z-test.

B. Experimental Baselines

Before feeding them to the classifier model, all the features extracted from our SnS are normalized as follows:

$$\hat{f}_{c,r} = \frac{f_{c,r} - Min(\mathbf{F}_c)}{Max(\mathbf{F}_c) - Min(\mathbf{F}_c)}, \quad (12)$$

where $f_{c,r}$ is the original c-th feature property for the r-th segment, $\mathbf{F}_c$ is the c-th feature vector which includes the properties for all of the segments. Thus, the normalized feature property $\hat{f}_{c,r}$ will be limited into [0,1].

TABLE IV
GRIDS OF MAIN PARAMETERS FOR EACH LEARNING MODEL.

Models	Main Parameters Setting
K-NN	*K*: 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100; distance metrics: 'euclidean', 'cityblock', 'chebychev', 'correlation', 'cosine', 'hamming', 'jaccard', 'minkowski', 'seuclidean', 'spearman'
LDA	discriminant type: linear; gamma: 0:0.05:1.00
SVM	'linear kernel', 'polynomial', 'radial basis function', 'sigmoid'; *c*-value: 10^{-5}, 10^{-4},..., 10^{4}, 10^{5}
RF	number of trees: 2^1, 2^2,..., 2^9, 2^{10}; fraction for the treebagger: 0.1:0.1:1.00
FNN	one hidden layer; size: 2^1, 2^2,..., 2^9, 2^{10}
ELM	activation functions: 'signmoidal', 'sine', 'hardlim', 'tribas', 'radbas'; number of hidden neurons: 2^1, 2^2,..., 2^{14}
KELM	kernels: 'radial basis function', 'linear', 'polynomial', 'wavelet'; regularization coefficients: 10^{-5}, 10^{-4},..., 10^{4}, 10^{5}

In order to evaluate the performance of our method, we apply the unweighted average recall (UAR), defined as:

$$UAR = \frac{\sum_{class=1}^{N_{MC}} N_{class,correct}/N_{class,all}}{N_{MC}} \times 100\%, \quad (13)$$

where $N_{class,correct}$, and $N_{class,all}$ are the number of correctly recognized segments, and all segments in one certain $class$, respectively. N_{MC} is the total number of classes.

The UAR baselines of the different combinations of classifiers and feature sets are shown in Table V. We found that *MFCCs* within a *K*-NN classifier (*K*: 30, distance metrics: 'correlation') achieve the best recognition rate of 76 % ($p < 0.001$). Of the nine feature sets, the novel wavelet-based *WEF* performs best with a mean UAR of 58 % among all classifiers used. *MFCCs* score second best (mean UAR of 56 %), followed by *SER* (UAR of 55 %). The performance of *WEF*, and *MFCCs* is significantly better compared to the remaining six feature sets ($p < 0.05$). On the other hand, *Crest Factor*, *F0*, and PR_{800} did not show a good performance in our study.

C. Feature Selection

We use the ReliefF algorithm as described in Section II-D for the feature selection step. The different feature sets separately, as well as a complete combination of all features from each feature set (*ALL*) are fed into the ReliefF process at *Rank Ratio* settings from 0.05 to 1.00 with increments of 0.05 to find the best-performing combined subset. The best results of the features selected by ReliefF are shown in TableVI.

We find that, except for the *Crest Factor* which remains at 34 % UAR, ReliefF can improve the mean performance of each feature set among different classifier models. In particular, for the combination of all feature sets, the mean performance significantly improves from 46 % to 68 % ($p < 0.001$). For *SFF* (45 % to 62 %), *MFCCs* (56 % to 68 %), and *WEF* (58 % to 66 %), the improvement after the feature selection step is also significant ($p < 0.001$, $p < 0.001$, and $p < 0.005$ respectively). The standard deviations among different classifiers in each of the feature sets decrease after the ReliefF step,

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

TABLE V
UAR ([%]) ACHIEVED WITH DIFFERENT FEATURE SETS AND CLASSIFIERS WITHOUT FEATURE SELECTION.

	K-NN	LDA	SVM	RF	FNN	ELM	KELM	Mean	Std. dev.
Crest Factor	37	36	30	32	38	37	29	34	±3.7
F0	32	37	34	27	36	31	32	33	±3.4
Formants	50	51	57	40	56	42	52	50	±6.5
SFF	60	59	25	66	55	25	25	45	±19.0
PR_{800}	35	39	36	35	39	36	41	37	±2.4
SER	62	56	51	66	50	44	53	55	±7.5
MFCCs	**76**	50	61	57	42	53	55	56	±10.5
EMDF	47	56	25	57	41	25	25	39	±14.5
WEF	59	53	53	59	56	59	64	58	±3.9
ALL	62	60	25	64	64	25	25	46	±20.1
Mean	52	50	40	50	48	38	40	–	–
Std. dev.	±14.3	±9.1	±14.3	±15.1	±9.7	±11.9	±14.8	–	–

which means that the selected subset of features is more robust and contains more useful information of the SnS compared to the original set. Finally, the best classification performance is achieved by a combination of all feature sets within a Random Forest classifier (UAR of 78 %). This reduced feature set has a dimension of 374, only 12.3 % of the original feature set.

Fig. 4 provides an insight of the weight contribution of different types of features for the classification of SnS. We found that, of all feature sets, *WEF* (containing *WPTE* and *WTE*), contributes most, followed by *MFCCs*, and *SFF*. This is not surprising given the high dimension compared to the other feature sets. As for functionals, *mean*, *max*, and *min* values of the frame-based LLDs contribute most, followed by the *bias* of the linear regression estimation. In addition, we illustrate the detailed weight contribution of *MFCCs*, and *WTE*. For *WTE*, we compare the contribution of the subsets by level of decomposition. It is shown that, the level-1 and level-2 decomposed components contribute most. In *MFCCs*, the *MFCCs*-7 and *MFCCs*-2 are best.

Fig. 5 shows the confusion matrix of the combination of all feature sets using a Random Forest classifier after ReliefF feature selection. We can see that, among the four classes of SnS, Type O (the oropharyngeal area), and Type E (the epiglottis) are the two most easily wrongly classified as Type E, and Type V (the level of the velum), respectively. In our experiments, the best trained classifier has the highest recognition accuracy for Type T (around 90 %), and the lowest accuracy for Type O (around 64 %).

IV. DISCUSSION

In this work, we systematically compare frequently used acoustic features for their performance on the classification of snore sounds based on the VOTE model. In our experiments, we can achieve a UAR of 78 % with the best combination of features and classifier. When performed by human experts, the interrater reliability of DISE classifications is up to 86 % [53]. This is a benchmark for us which has not yet been entirely achieved by our model. We believe that a major limitation is the small number of independent subjects in the database, and we aim to improve our results based on more data.

For our data set of snore sounds and with the experiments described, *MFCCs*, and *WEF* have shown to be the best suited feature sets (mean UAR at 68 % and 66 %, respectively) across different classifiers. As a sophisticated indicator in speech recognition, *MFCCs* can be regarded to represent the airway transfer function. In our previous studies on classification of different snore related sounds from overnight audio recordings, *MFCCs* also proved to be quite efficient. Thus, *MFCCs* tend to play an important role SnS classification.

The proposed novel feature set, *WEF*, is based on the wavelet transform theory, which is capable to give a multi-resolution analysis of non-stationary signals. In the early works by Matsiki et al. [36], Continuous Wavelet Transform (CWT) is used to analyse the spectrum energy distribution changes before, during, and after apneic events based on snoring sounds. However, they did not propose a feature extraction method and their classification is not based on machine learning techniques. Furthermore, the number of subjects they investigated is small with only seven in total. Ng et al. [54] used wavelet polyspectral techniques to generate novel features to distinguish apneic from non-apneic snoring. Their results in [54] showed that wavelet based features outperform the conventional spectral peak frequency. In our study, the *WEF* feature set, combining both the wavelet packet transform and the wavelet transform techniques, achieves an excellent performance of mean UAR of 66 % among all classifiers used after the feature selection step. Further, it contributes most (69 %) in the best-performing feature set.

Ng et al. [55] describe that formants are representative of the physical frequency transfer function of the upper airway and can be good indicators to classify apneic and non-apneic snorers. The performance of the *Formants* feature set in our study is better than that of other features analysed in earlier studies, i. e., *Crest Factor*, *F0*, and PR_{800} ($p < 0.001$). This can be another indicator that the airway transfer function is a suitable indicator to distinguish different forms of snoring.

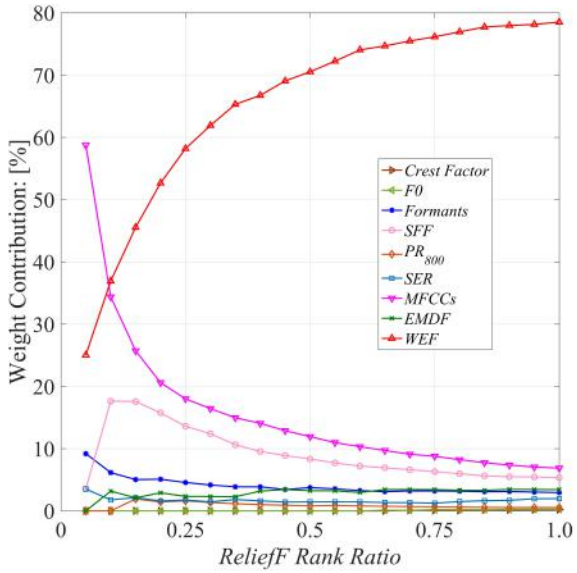
EMD is based on a different signal analysis method than the FFT. Specifically, it is suitable for non-stationary signals [33]. In this study, *EMDF* did not prevail as a feature set. A possible reason for this moderate performance is that we based the classification task on relatively short fractions of a snoring event in order to increase the number of available segments for model learning. Therefore, the non-stationary characteristics of the underlying snore event could not be considered in full. A better performance can be expected when using this feature set on complete snoring events.

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

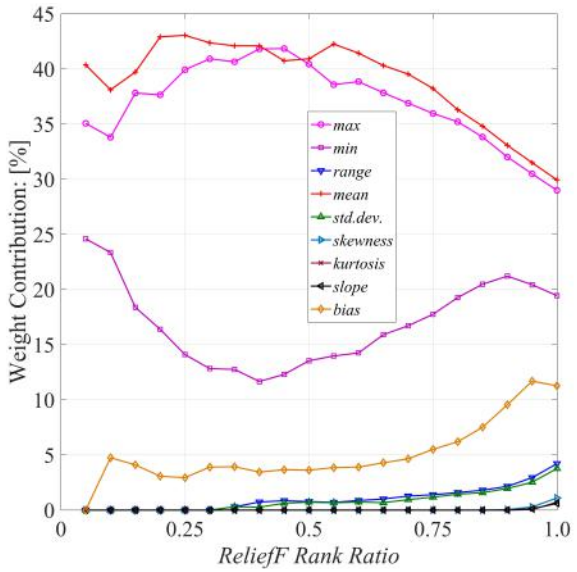
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

TABLE VI
UAR ([%]) ACHIEVED WITH DIFFERENT FEATURE SETS AND CLASSIFIERS AFTER THE FEATURE SELECTION STEP.

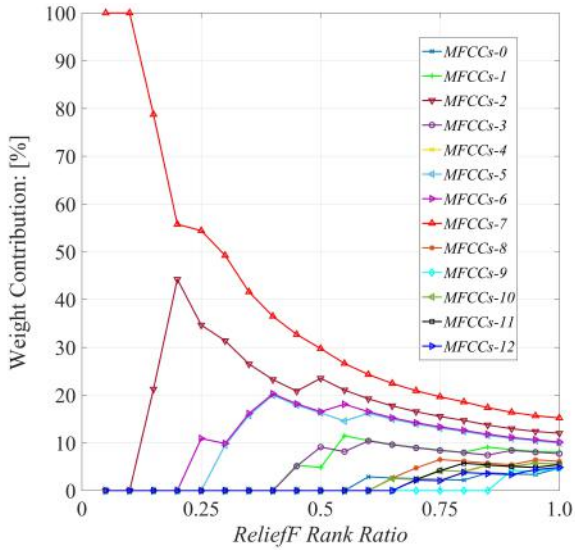
	K-NN	LDA	SVM	RF	FNN	ELM	KELM	Mean	Std. dev.
Crest Factor	35	36	31	36	39	30	30	34	±3.5
F0	32	38	34	34	39	33	32	35	±2.8
Formants	54	50	55	50	57	50	57	53	±3.3
SFF	62	71	65	72	62	54	49	62	±8.4
PR_{800}	35	39	40	36	42	42	42	39	±2.9
SER	61	58	58	66	58	52	59	59	±4.2
MFCCs	76	64	62	70	73	66	64	68	±5.2
EMDF	52	55	60	58	56	47	54	55	±4.2
WEF	68	69	69	67	64	59	68	66	±3.6
ALL	74	73	74	**78**	71	45	58	68	±11.8
Mean	55	55	55	57	56	48	51	–	–
Std. dev.	±16.3	±14.1	±14.8	±16.6	±12.4	±11.0	±12.9	–	–



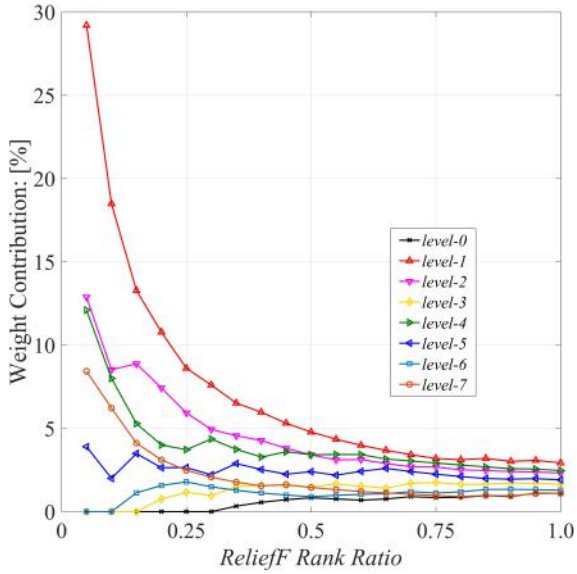
(a) Detailed weight contribution of all feature sets



(b) Detailed weight contribution of functionals



(c) Detailed weight contribution of MFCCs



(d) Detailed weight contribution of *WTE*

Fig. 4. Weight contribution ([%]) analysis by the ReliefF algorithm with different *Rank Ratio*s.

Future works can be done mainly in the following three areas: Firstly, more potentially useful acoustic features can be

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

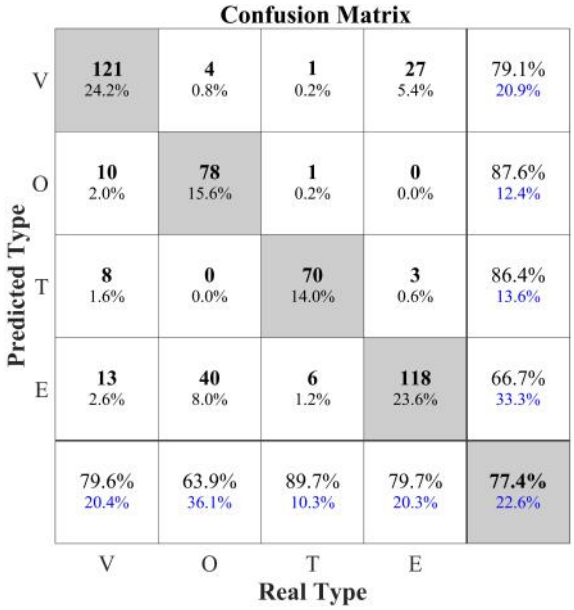


Fig. 5. The confusion Matrix of the best-performing combination of feature set and classifier.

tested, specifically, psychoacoustic characteristics (e. g., loudness, sharpness, roughness, fluctuation strength), and higher order statistical model based features (e. g., bispectrum), which have been studied in [56], [57]. Some fundamental work to explore the relationship between feature properties and the anatomical changes in the upper airway can help to better understand the SnS generation mechanisms. Also, using complete snore events, rather than segments of snore sounds, as a basis for feature extraction can reveal additional useful information on the different snoring classes, as their non-stationary characteristics will show more clearly. A limitation of our work is the relatively small number of snoring subjects in the database used. Although the total number of snore segments is sufficiently large to apply machine learning methods, they stem from comparably few different subjects. Last but not least, the feature selection phase at this stage of our method is based on an empirical parameter setting process (*Rank Ratio* is set using a step interval selection process) rather than on an automatic method without human involvement. Excluding the 'human touch' on the parameter setting process will be important for future baseline improvement and practical product development.

V. Conclusion

For the first time, we comprehensively investigated various acoustic feature sets and classifiers for the task of classifying snore sounds according to their excitation locations based on the VOTE model. Even with a relatively small data set, we can achieve a good classification performance with selected feature sets independent of subjects. The results show that multi-feature analysis is a promising means to help identifying the anatomical mechanisms of snore sound generation in individual subjects.

Acknowledgment

The authors would like to thank all the colleagues involved in the collection of the labelled VOTE snore sounds data. This work is supported by the China Scholarship Council (CSC), the European Union's Seventh Framework, and Horizon 2020 under grant agreements No. 338164 (ERC Starting Grant iHEARu), and No. 645378 (ARIA-VALUSPA).

References

[1] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," *American Journal of Epidemiology*, vol. 177, no. 9, pp. 1006–1014, 2013.

[2] P. J. Strollo Jr and R. M. Rogers, "Obstructive sleep apnea," *New England Journal of Medicine*, vol. 334, no. 2, pp. 99–104, 1996.

[3] P. Smith, D. Hudgel, L. Olson, M. Partinen, D. Rapoport, C. Rosen, J. Skatrud, R. Waldhorn, P. Westbrook, and T. Young, "Indications and standards for use of nasal continuous positive airway pressure (cpap) in sleep-apnea syndromes," *American Journal of Respiratory and Critical Care Medicine*, vol. 150, no. 6, pp. 1738–1745, 1994.

[4] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New England Journal of Medicine*, vol. 328, no. 17, pp. 1230–1235, 1993.

[5] B. Mokhlesi, S. A. Ham, and D. Gozal, "The effect of sex and age on the comorbidity burden of osa: an observational analysis from a large nationwide us health claims database," *European Respiratory Journal*, vol. 47, no. 5, pp. ERJ–01 618, 2016.

[6] M. S. Aldrich, *Sleep Medicine*. Transaction Publishers, 1999.

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

[8] A. Roebuck, V. Monasterio, E. Gederi, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G. Clifford, "A review of signals used in sleep analysis," *Physiological Measurement*, vol. 35, no. 1, pp. R1–R57, 2014.

[9] K. K. Li, "Surgical therapy for adult obstructive sleep apnea," *Sleep Medicine Reviews*, vol. 9, no. 3, pp. 201–209, 2005.

[10] H.-C. Lin, M. Friedman, H.-W. Chang, and B. Gurpinar, "The efficacy of multilevel surgery of the upper airway in adults with obstructive sleep apnea/hypopnea syndrome," *The Laryngoscope*, vol. 118, no. 5, pp. 902–908, 2008.

[11] M. R. El Badawey, G. McKee, H. Marshall, N. Heggie, and J. A. Wilson, "Predictive value of sleep nasendoscopy in the management of habitual snorers," *Annals of Otology, Rhinology & Laryngology*, vol. 112, no. 1, pp. 40–44, 2003.

[12] K. Behbehani, F.-C. Yen, J. R. Burk, E. A. Lucas, and J. R. Axe, "Automatic control of airway pressure for treatment of obstructive sleep apnea," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 10, pp. 1007–1016, 1995.

[13] H. Demin, Y. Jingying, W. J. Y. Qingwen, L. Yuhua, and W. Jiangyong, "Determining the site of airway obstruction in obstructive sleep apnea with airway pressure measurements during sleep," *The Laryngoscope*, vol. 112, no. 11, pp. 2081–2085, 2002.

[14] M. Reda, G. J. Gibson, and J. A. Wilson, "Pharyngoesophageal pressure monitoring in sleep apnea syndrome," *Otolaryngology–Head and Neck Surgery*, vol. 125, no. 4, pp. 324–331, 2001.

[15] B. A. Stuck and J. T. Maurer, "Airway evaluation in obstructive sleep apnea," *Sleep Medicine Reviews*, vol. 12, no. 6, pp. 411–436, 2008.

[16] S. Miyazaki, Y. Itasaka, K. Ishikawa, and K. Togawa, "Acoustic analysis of snoring and the site of airway obstruction in sleep related respiratory disorders," *Acta Oto-Laryngologica*, vol. 118, no. 537, pp. 47–51, 1998.

[17] P. Hill, B. Lee, J. Osborne, and E. Osman, "Palatal snoring identified by acoustic crest factor analysis," *Physiological Measurement*, vol. 20, no. 2, pp. 167–174, 1999.

[18] S. Agrawal, P. Stone, K. McGuinness, J. Morris, and A. Camilleri, "Sound frequency analysis and the site of snoring in natural and induced sleep," *Clinical Otolaryngology & Allied Sciences*, vol. 27, no. 3, pp. 162–166, 2002.

[19] R. J. Beeton, I. Wells, P. Ebden, H. Whittet, and J. Clarke, "Snore site discrimination using statistical moments of free field snoring sounds recorded during sleep nasendoscopy," *Physiological Measurement*, vol. 28, no. 10, pp. 1225–1236, 2007.

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

[20] C. Janott, W. Pirsig, and C. Heiser, "Akustische analyse von schnarchgeräuschen," *Somnologie-Schlafforschung und Schlafmedizin*, vol. 18, no. 2, pp. 87–95, 2014.

[21] E. J. Kezirian, W. Hohenhorst, and N. de Vries, "Drug-induced sleep endoscopy: the vote classification," *European Archives of Oto-Rhino-Laryngology*, vol. 268, no. 8, pp. 1233–1236, 2011.

[22] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins, "Multi-feature snore sound analysis in obstructive sleep apnea–hypopnea syndrome," *Physiological Measurement*, vol. 32, no. 1, pp. 83–97, 2011.

[23] K. Qian, Y. Fang, Z. Xu, and H. Xu, "Comparison of two acoustic features for classification of different snore signals," *Chinese Journal of Electron Devices*, vol. 36, no. 4, pp. 455–459, 2013.

[24] X. Sun, "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio," in *Proc. of the IEEE ICASSP*, vol. 1. Orlando, Florida, USA: IEEE, 2002, pp. 333–336.

[25] J. R. Deller Jr, J. G. Proakis, and J. H. Hansen, *Discrete Time Processing of Speech Signals*. Prentice Hall PTR, 1993.

[26] T. Emoto, U. Abeyratne, M. Akutagawa, H. Nagashino, and Y. Kinouchi, "Feature extraction for snore sound via neural network processing," in *Proc. of the IEEE EMBS Annual International Conference*. Lyon, France: IEEE, 2007, pp. 5477–5480.

[27] K. Qian, Z. Xu, H. Xu, and B. P. Ng, "Automatic detection of inspiration related snoring signals from original audio recording," in *Proc. of the IEEE ChinaSIP*. Xi'an, China: IEEE, 2014, pp. 95–99.

[28] W. Whitelaw, "Characteristics of the snoring noise in patients with and without occlusive sleep apnea," *American Review of Respiratory Disease*, vol. 147, pp. 635–644, 1993.

[29] 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.

[30] A. Azarbarzin and Z. Moussavi, "Automatic and unsupervised snore sound extraction from respiratory sound signals," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1156–1162, 2011.

[31] L. R. Rabiner and R. W. Schafer, "Theory and application of digital speech processing," *Preliminary Edition*, 2009.

[32] 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.

[33] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971. The Royal Society, 1998, pp. 903–995.

[34] Y. Yu, C. Junsheng *et al.*, "A roller bearing fault diagnosis method based on emd energy entropy and ann," *Journal of Sound and Vibration*, vol. 294, no. 1, pp. 269–277, 2006.

[35] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[36] D. Matsiki, X. Deligianni, E. Vlachogianni-Daskalopoulou, and L. J. Hadjileontiadis, "Wavelet-based analysis of nocturnal snoring in apneic patients undergoing polysomnography," in *Proc. of the IEEE EMBS Annual International Conference*. Lyon, France: IEEE, 2007, pp. 1912–1915.

[37] A. K. Ng, T. San Koh, K. Puvanendran, and U. R. Abeyratne, "Snore signal enhancement and activity detection via translation-invariant wavelet transform," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 10, pp. 2332–2342, 2008.

[38] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 121–131, 2011.

[39] R. R. Coifman, Y. Meyer, and V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and their Applications*. Sudbury: MA: Jones and Barlett, 1992, pp. 153–178.

[40] K. B. Englehart, "Signal representation for classification of the transient myoelectric signal," Ph.D. dissertation, University of New Brunswick, Department of Electrical & Computer Engineering, 1998.

[41] K. Qian, C. Janott, Z. Zhang, C. Heiser, and B. Schuller, "Wavelet features for classification of vote snore sounds," in *Proc. of the IEEE ICASSP*. Shanghai, China: IEEE, 2016, pp. 221–225.

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

[43] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[44] I. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3–31, 2000.

[45] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 2013.

[46] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[47] T. G. Dietterich, "Ensemble learning," *The Handbook of Brain Theory and Neural Networks*, vol. 2, pp. 110–125, 2002.

[48] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[49] G.-B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.

[50] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[51] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Proc. of the AAAI*, vol. 2, San Jose, California, USA, 1992, pp. 129–134.

[52] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with relieff," *Applied Intelligence*, vol. 7, no. 1, pp. 39–55, 1997.

[53] E. J. Kezirian, D. P. White, A. Malhotra, W. Ma, C. E. McCulloch, and A. N. Goldberg, "Interrater reliability of drug-induced sleep endoscopy," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 136, no. 4, pp. 393–397, 2010.

[54] A. K. Ng, T. San Koh, U. R. Abeyratne, and K. Puvanendran, "Investigation of obstructive sleep apnea using nonlinear mode interactions in nonstationary snore signals," *Annals of Biomedical Engineering*, vol. 37, no. 9, pp. 1796–1806, 2009.

[55] A. K. Ng, T. San Koh, E. Baey, T. H. Lee, U. R. Abeyratne, and K. Puvanendran, "Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?" *Sleep Medicine*, vol. 9, no. 8, pp. 894–898, 2008.

[56] M. Herzog, S. Plößl, A. Glien, B. Herzog, C. Rohrmeier, T. Kühnel, S. Plontke, and P. Kellner, "Evaluation of acoustic characteristics of snoring sounds obtained during drug-induced sleep endoscopy," *Sleep and Breathing*, pp. 1–9, 2014.

[57] A. K. Ng, K. Y. Wong, C. H. Tan, and T. S. Koh, "Bispectral analysis of snore signals for obstructive sleep apnea detection," in *Proc. of the IEEE EMBS Annual International Conference*. Lyon, France: IEEE, 2007, pp. 6195–6198.

Kun Qian (S'14) received his master degree from Nanjing University of Science and Technology (NUST) in P. R. China, 2014. From November 2013 to April 2014, he was sponsored by a full fellowship of NUST for Master Program students (10 among 4 600) to Nanyang Technological University (NTU), Singapore as a Research Assistant at Information Systems Research Laboratory. Now he is working on his PhD degree at Technische Universität München (TUM), Germany, and a Visiting Doctoral Student at Matsuoka Lab in Tokyo Institute of Technology, Japan, under the project of fast large amount of audio data classification by high performance computing system. His research interests include: signal processing, machine learning, biomedical engineering, and high performance computing system based machine learning.

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

Christoph Janott (S'15) graduated as electrical engineer (Dipl. -Ing.) at Technische Universität Berlin (TUB) in Germany, 1998. Since that time, he has held several leadership positions in product management and marketing in different technology companies. Since 2011, he coaches founders and young entrepreneurs in the field of medical devices in marketing and business development. He is co-author of several publications on speech intelligibility in classrooms and has co-invented several patents. Currently, he is a doctoral candidate at the Institute of Medical Engineering at Technische Universität München (TUM), Germany.

Vedhas Pandit (S'11) received his master degree from Arizona State University (ASU) in USA, 2010 in Electronic and mixed signal circuit design (EECE) with his thesis on mathematical modeling of a-Si:H SOI transistors. After working for Intel as a Graphics Hardware Engineer, he worked as a researcher at Indian Institute of Technology Bombay (IITB) developing tools for automated music information retrieval, specifically for recurring pattern /melodic motif detection and identification, pattern classification, pattern variability characterization (e. g. non-uniform time warping, ornamentations) and investigating relevance of each to Hindustani classical music tradition. Since February 2015, he has been working on his PhD degree at University of Passau, Passau, Germany. His research interests include: music information retrieval, speech and virtual instrument synthesis, deep learning strategies in machine learning, and biomedical signal processing.

Zixing Zhang (M'15) received his master degree in physical electronics from Beijing University of Posts and Telecommunications, China, 2010, and his PhD degree in engineering from the Institute for Human-Machine Communication at Technische Universität München (TUM), Germany, 2015. He is currently a postdoctoral researcher at the University of Passau, Germany. He has authored about thirty publications in peer-reviewed journals and conference proceedings. His research interests mainly lie in semi-supervised learning, active learning, multi-task learning, and deep learning in the application of computational paralinguistics, and automatic speech recognition.

Clemens Heiser graduated from medical school at University of Heidelberg in Germany 2007. He began his residency in Otolaryngology (head and neck surgery) at Klinikum Mannheim (University of Heidelberg) and completed it at Klinikum Rechts der Isar (Technische Universität München (TUM)) in 2011. He also completed his fellowship training in sleep medicine at Klinikum Mannheim and is a board certified member of the German Sleep Society. Since 2013 he is the head of the Sleep Departement at Klinikum rechts der Isar. His research and specialty is in the surgical treatment of snoring and Obstructive Sleep Apnea (OSA). Dr. Heiser is involved in developing surgical techniques and various alternative treatments for patients with OSA. He is co-author of the German guidelines in the treatment of snoring and OSA.

Winfried Hohenhorst graduated from medical school at University of Essen in Germany 1985. He concluded his residency in Anaesthesiology, Otolaryngology and Plastic Surgery at Alfried Krupp Hospital in Essen in 1992. He is a board certified member of the German Sleep Society, and a specialist for voice and speech disorders, allergology and pastic operations. Since 2012 he is the head of the clinic for ENT, plastic surgery and operative sleep medicine at Alfried Krupp Hospital in Essen. His research activities are focused on the diagnosis of sleep related breathing disorders by means of sleep endoscopy, as well as the surgical treatment of snoring and Obstructive Sleep Apnea (OSA). He is co-author of the German guidelines for the treatment of snoring and OSA.

Michael Herzog graduated from medical school at University of Würzburg, Germany, 1998. He finished his residency in Otorhinolaryngology at the ENT department of the University of Würzburg in 2005. 2005 until 2015 he worked as a senior consultant at the ENT department of the university of Greifswald, Germany as well as Halle (Saale), Germany. Since 2015 he is head of the Department of Otorhinolaryngology, Head and Neck Surgery at Carl-Thiem-Klinikum, Cottbus, Germany. He is a board certified member of the German Sleep Society and a specialist for oncological, reconstructive and plastic surgery. His research activities are focused on the diagnosis and surgical treatment of snoring and sleep apnea as well as the research in acoustic analysis of snoring sounds. He is co-author of the German guidelines for the treatment of snoring and OSA..

Werner Hemmert (M'00–SM'08) explores the principles of information processing in the auditory system and novel approaches for the advancement of neuroprostheses. His research is characterized by a combination of theoretical concepts and experiments and he relies on close collaboration with workgroups from the fields of biology, medicine and industry. After studying Electrical Engineering and Computer Engineering at the Technische Universität München (TUM), he contimued his research career at the Tübingen Hearing Research Center (1991–1998), where he studied the cellular micromechanics of the inner ear. He submitted and received his doctorate at Ruhr-Universität Bochum in 1997. After that, he conducted research at the Massachusetts Institute of Technology (1998–2000), the IBM Research Laboratories in Zürich (2000–2001) and at Infineon Technologies, Corporate Research (2001–2007). Since 2007 he is professor for Bio-Inspired Information Processing at TUM. Prof. Hemmert is, among others, member of the German Acoustical Society (DEGA), the Association for Research in Otolaryngology and the Graduate School of Systemic Neurosciences and the Alexander von Humboldt Foundation.

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2016.2619675, IEEE Transactions on Biomedical Engineering

Björn Schuller (M'06–SM'15) received his diploma in 1999, his doctoral degree for his study on automatic speech and emotion recognition in 2006, and his habilitation and Adjunct Teaching Professorship in the subject area of signal processing and machine intelligence in 2012, all in electrical engineering and information technology from Technische Universität München (TUM), Germany. He is a tenured Full Professor heading the Chair of Complex Systems Engineering at the University of Passau, Germany, and a Reader (Associate Professor) in Machine Learning in the in the Department of Computing at the Imperial College London in London, UK. Dr. Schuller is elected member of the IEEE Speech and Language Processing Technical Committee, Editor in Chief of the IEEE Transactions on Affective Computing, and senior member of the IEEE, ACM and ISCA and (co-)authored 5 books and more than 500 publications in peer reviewed books, journals, and conference proceedings leading to more than 10 000 citations.

0018-9294 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.