

WAVELET FEATURES FOR CLASSIFICATION OF VOTE SNORE SOUNDS

Kun Qian¹, Christoph Janott², Zixing Zhang⁵, Clemens Heiser³, Björn Schuller^{4,5}

¹MISP group, MMK, Technische Universität München, Germany

²Institute for Medical Engineering, Technische Universität München, Germany

³Department of Otorhinolaryngology/Head and Neck Surgery,
Klinikum rechts der Isar, Technische Universität München, Germany

⁴Machine Learning Group, Imperial College London, UK

⁵Chair of Complex & Intelligent Systems, University of Passau, Germany

andykun.qian@tum.de, schuller@ieee.org

ABSTRACT

Location and form of the upper airway obstruction is essential for a targeted therapy of obstructive sleep apnea (OSA). Utilizing snore sounds (SnS) to reveal the pathological characters of OSA patients has been the subject of scientific research for several decades. Fewer studies exist on the evaluation of SnS to identify the corresponding obstruction types in the upper airway. In this study, we propose a novel feature set based on wavelet transform with a support vector machine classifier to discriminate VOTE (velum, oropharyngeal lateral walls, tongue base and epiglottis) snore sounds labelled during drug-induced sleep endoscopy (DISE). Based on snore sound data collected from 24 snoring subjects, processed by a subject-independent 2-fold cross validation experiment, we can show that our wavelet features outperform the frequently-used acoustic features (formants, MFCC, power ratio, crest factor, fundamental frequency) at an WAR (weighted average recall) of 78.2% and an UAR (unweighted average recall) of 71.2%, with an enhancement ranging from 5.1% to 24.4% and 12.2% to 46.4% in WAR and UAR, respectively.

Index Terms— Snore Sounds, Obstructive Sleep Apnea, Upper Airway, Wavelet Transform

1. INTRODUCTION

Obstructive sleep apnea (OSA) is a chronic disease which affects 13% of men and 6% of women in the USA [1]. Symptoms of OSA can include, among others, daytime sleepiness, and morning headache [2]. It is associated with reduced quality of life and an independent risk factor for hypertension [3], myocardial infarction [4] and stroke [5]. OSA is defined as a syndrome with cessation of airflow for more than 10 seconds (5 or more episodes per hour in sleep), usually associated with a decrease of oxyhemoglobin saturation [6]. Loud snoring is a typical symptom of OSA, reported in more than 80% of OSA patients [7]. With the state-of-the-art theories and techniques in acoustic analysis and modelling, there are many pioneers [8] studying how to use snore sounds (SnS) to distinguish OSA patients and primary snorers, with promising results. The precise analysis of the underlying individual anatomical mechanisms of snoring and OSA can lead to a targeted and less invasive surgical approach. Today, drug induced sleep endoscopy (DISE) is increasingly used to identify location and form of vibrations and obstructions [9] in order to evaluate the obstruction level and pattern. However, DISE is

time consuming, costly, and places patients under strain. Determination of the excitation location within the upper airways by means of acoustic analysis of the snoring noise could be easier for doctor and patient. However, there are few studies on the subject of determining the obstruction or vibration sites within the upper airway based on acoustic features. Miyazaki et al. studied the relationship of the fundamental frequency of SnS and the obstruction site, distinguishing between soft palate, tonsils/tongue base, combined type (both palate and tonsils/tongue base), and the larynx [10]. The crest factor, a simple but practical acoustic character, had been demonstrated to be efficient to distinguish palatal and non-palatal snorers [11]. Agrawal et al. compared peak frequency, centre frequency and power ratio variations of different snores and the distinction between natural sleep and induced sleep [12]. Beeton et al. indicated that with a combination of a 2-means clustering method and the statistical dimensionless moment coefficients of skewness and kurtosis, they could discriminate palatal and non-palatal snoring [13].

In the above mentioned studies, basic acoustic features have been evaluated for their suitability and performance to distinguish between different types of snoring. But the models based on machine learning and signal processing have not yet been described for this purpose. It must be noted that the ‘snore site’ and the ‘obstruction site’ in the upper airway are two different definitions, which may or may not coincide in individual patients. In our work, we exclusively focus on the determination of the site of vibration as a cause for the generation of snore sounds. Our definition of the snore site is in line with the ‘VOTE’ concept, introduced by Kezirian et al. for the use in DISE evaluation [14]. Based on this classification we distinguish between the velum, the oropharyngeal area, the tongue Base, and the epiglottis level to define the classes of snorers.

In this study, we present a wavelet transformation based feature set with a support vector machine (SVM) for classification of VOTE snore sounds. This wavelet transformation based method originated from the work by Khushaba et al. [15], which has been demonstrated to be effective in electroencephalogram (EEG), electrooculogram (EOG), and electrocardiogram (ECG) signal processing. We elaborate on selecting a suitable ‘wavelet type’, namely, the wavelet function for extraction of descriptors from the SnS. We compare our proposed feature set with some frequently-used acoustic features (e. g., formants, MFCC, power ratio, crest factor, and fundamental frequency). The results prove that wavelet features can outperform other frequently-used feature sets.

Table 1. Anthropological information (age, body mass index (BMI) and Apnea/Hypopnea Index (AHI)) of the patients.

	mean	std	range
Age [years]	46.2	13.2	26 – 72
BMI [kg/m^2]	26.8	2.9	18.9 – 31.5
AHI [events/h]	20.4	10.7	6.2 – 45.6

Table 2. Instance numbers for each data group and the distribution of our V (the velum), O (the oropharyngeal area), T (the tongue Base), and E (the epiglottis level) Type SnS data.

	V	O	T	E	All
Group 1	376	132	18	125	651
Group 2	434	111	46	141	732
Total	810	234	64	266	1383

2. MATERIALS AND METHODS

2.1. Patients and Data

This study is approved by the ethic committee of Klinikum rechts der Isar, Technische Universität München, Germany. We collected SnS data from 24 subjects diagnosed with primary snoring or OSA through a polysomnography (PSG) beforehand. The anthropological information of patients, namely the age, body mass index (BMI), and the Apnea/Hypopnea Index (AHI) are listed in Table 1. DISE was performed in all patients in order to determine adequate surgical intervention measures. The DISE video was recorded using a flexible nasopharyngoscope; audio information was recorded in parallel using a headset microphone and synchronously stored in the same file (MP4). In the video and audio recordings, the locations of sound generation were categorised based on the VOTE classification by an ENT (ear, nose, and throat) expert. 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 total, we used 117 snoring episodes (time lasts from 0.31 to 2.17 s with an average of 1.24 s), which include 66, 20, 10, and 21 episodes of V, O, T, E types of SnS, respectively. The audio files have a sampling frequency of 16 kHz with 16-bit resolution. We segmented the episodes into single instances for further feature extraction and machine processing. Every instance has a length of 200 ms and neighbouring instances have an overlap of 50%. To make full use of our limited data, we performed a 2-fold cross validation-based subject-independent evaluation. We randomly separated the 24 patient’s data into two groups, i. e., each group has the SnS instances from 12 patients. The number of instances for each group are shown in Table 2.

2.2. Wavelet Features

In 2007, Matsiki et al. firstly studied the use of wavelet-based methods to analyse SnS of OSA patients [16]. They adopted the con-

tinuous wavelet transform (CWT) with a Morlet wavelet function to analyse the energy distribution of SnS before, during, and after an OSA event. They indicated that the energy distribution of a snoring episode directly following an apneic event is higher compared to a snoring episode before or during an apnea. However, they did not reveal whether wavelet-based features could reflect the pathological characters of the obstruction site in the upper airway. Figure 1 illustrates the signal waveforms and corresponding scalograms of the four types of SnS. The scalogram indicates the energy percentage for each wavelet coefficient (the wavelet function here is ‘ $db10$ ’¹). The scalogram is based on CWT, to analyse time and frequency distribution of SnS (cf. also [16]). The scalogram reveals the energy distribution of the wavelet transformed signal corresponding to different scales of the decomposition process. From the scalogram, one sees that, E type SnS shows more energy units at higher scales (around 20) compared to the other three types. This means that, in our set of samples, the SnS of type E contains more energy at a lower frequency band compared with the other three types. The scalogram also depicts that, the V type SnS has a higher energy content in the lower and higher scales, which is similar to the O type. The T type SnS shares more similarities with E type SnS both in the signal waveform in the time domain and the scalogram in the wavelet transform domain.

Khushaba et al. proposed a wavelet-packet-based feature-extraction algorithm and utilised it to classify EEG, EOG, and ECG signals [15], where results showed that the extracted wavelet-based features could achieve a classification accuracy of (95 % to 97 %) for different drowsiness levels. In this paper, we extend this work into SnS classification and explore the most suitable wavelet function for the feature set extraction. The wavelet packet transform (WPT) was introduced by Coifman et al. [17], which aims to reveal the links between multi-resolution approximations and wavelets. The WPT can be understood as a tree of subspaces, where $V_{0,0}$ is the root node. In general notation, the signal space $V_{j,k}$ (j is the level of the decomposition process, and k is the sub-band index) is decomposed into two orthogonal subspaces level by level: $V_{j+1,2k}$ and $V_{j+1,2k+1}$, namely, the ‘approximation space’ and the ‘detail space’ [18]. This decomposition process is done by dividing an orthogonal basis $\Phi_j(t - 2^j k)_{j,k \in \mathbb{Z}}$ from $V_{j,k}$ into two new orthogonal bases: $\Phi_{j+1}(t - 2^{j+1} k)_{j,k \in \mathbb{Z}}$ from $V_{j+1,2k}$ and $\Psi_{j+1}(t - 2^{j+1} k)_{j,k \in \mathbb{Z}}$ from $V_{j+1,2k+1}$, respectively, where, $\Phi_{j,k}(t)$ and $\Psi_{j,k}(t)$ are wavelet functions [19]. WPT can reveal the subband features from the given SnS data with the decomposition process, presented by Khushaba et al. [15]. A construction of normalised filter bank energy is defined as:

$$E_{V_{j,k}} = \sqrt{\frac{\sum_n (w_{j,k,n})^2}{N_k}}, \quad (1)$$

where $w_{j,k}$ represents the wavelet-packet transform coefficients evaluated from the signal at the subspace $V_{j,k}$, and N_k is the number of wavelet coefficients in the k -th subband; therefore, $E_{V_{j,k}}$ denotes the normalised bank filter energy in k -th subband with the j -th decomposition level. Furthermore, the subband energy percentage is defined as:

$$E_{V_j} = \frac{\sum_k \sum_n (w_{j,k,n})^2}{\sum_{j=1}^{J_{max}} \sum_k \sum_n (w_{j,k,n})^2}. \quad (2)$$

¹The wavelet function names mentioned here and on following are identical to the ones given in the Matlab Wavelet Toolbox (<http://www.mathworks.com/products/wavelet/>).

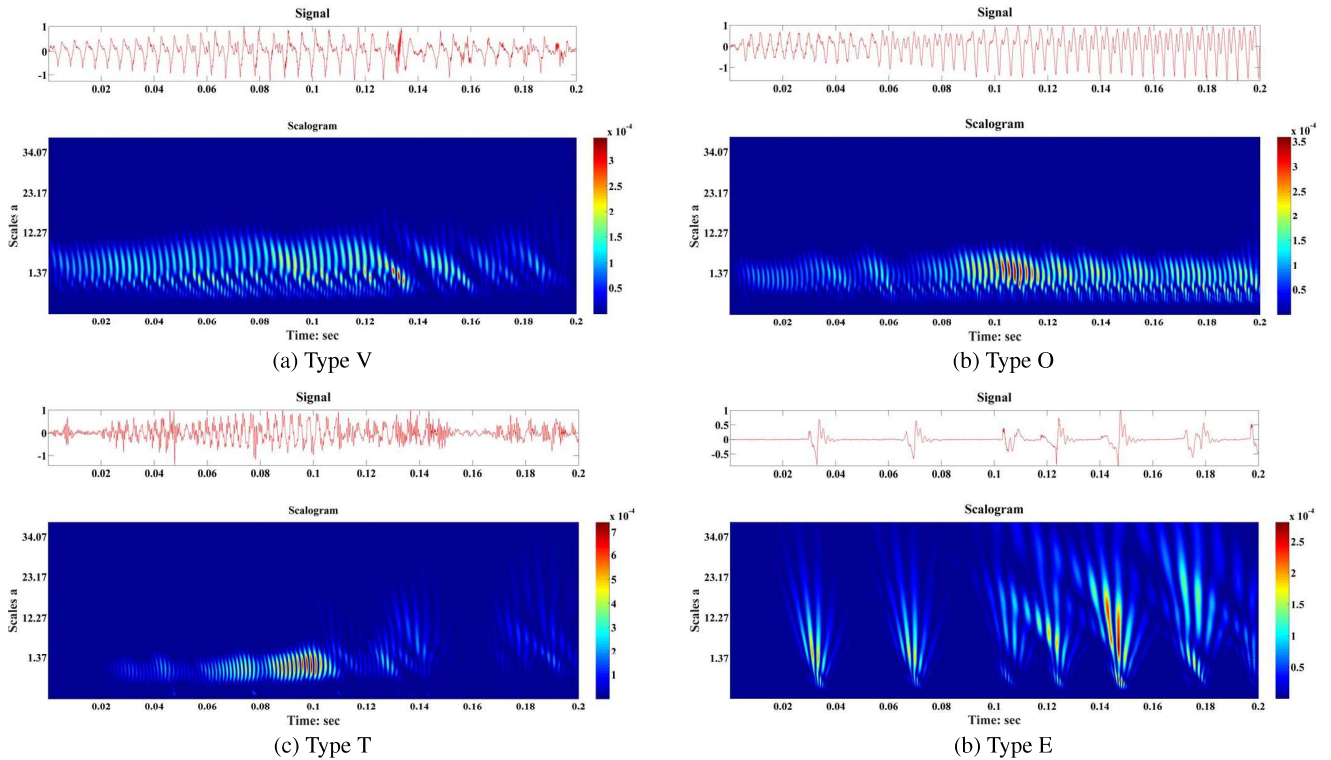


Fig. 1. Signal waveform and scalogram of VOTE typical SnS samples.

Subsequently, the variance, waveform length (the sum of the absolute differences), and the entropy are calculated from the base by Eq. (2). Additionally, all the energy-related features presented in Eq. (1) and Eq. (2) are modified with a logarithmic operator. From Eq. (1) we obtain $2^{J_{max}+1} - 1$ WPT energy related features and with Eq. (2) we extract $4 \times (J_{max} + 1)$ features, where J_{max} is the maximum level for WPT decomposition (in our study it is set to be 6; thus we obtain 155 descriptors). Motivated by the successful performance achieved by our large scale feature extraction toolkit, openSMILE [20], we apply 9 statistical functionals to the above low-level descriptors (LLDs). In our feature extraction algorithm, for each instance, the original SnS data are divided into frames (each frame contains 64 ms with a 50% overlap rate). Then, we calculate the statistical values from among these frames and obtain a complete attribute information per analysed instance. The detailed process of this technique is described in [20]. We illustrate the basic wavelet features and the statistical functions in Table 3. Among each of these, the *delta coefficients* (velocity), and *delta-delta coefficients* (acceleration) are computed; therefore, the total number of attributes per feature vector is $155 \times 9 \times 3 = 4185$.

3. RESULTS AND DISCUSSIONS

3.1. Experimental Setup

Due to the high popularity and performance of SVMs [21] in audio event analysis, we choose them as our classifier for the experiments to follow. The frequently-used and mature toolkit LIBSVM [22] is used, and here we chose an SVM with a linear kernel. All parameters are set to the default values, which could achieve the best

Table 3. Basic wavelet features (LLDs) and applied statistical functionals.

LLDs (155)	Statistical functionals (9)
$E_{V_{j,k}}$ (127)	max, min, mean,
E_{V_j} (7)	range, standard deviation,
Variance of E_{V_j} (7)	slope, bias (linear
Waveform length of E_{V_j} (7)	regression approximation)
Entropy of E_{V_j} (7)	skewness, kurtosis

performance in our experiments. To make full use of our limited database, we adopt a k -fold cross validation method (in this study, k is 2), which uses subject-independent (adopted by Alshaer et al. in recent OSA related study [23]) data groups as shown in Table 2. The evaluation value – the WAR (weighted average recall, i. e., the accuracy) and UAR (unweighted average recall, i. e., mean value of WAR for each class of testing instances) – are all the mean values achieved across the 2-fold experiments. We need to note that in the previous studies, WAR was normally used rather than UAR, which can risk to ignore the fact that SnS data are naturally imbalanced data. Therefore, we adopt both WAR and UAR together to evaluate the performance of our proposed method in this study. All the experiments and algorithms are implemented in the environment of Matlab R2014b by MathWorks®.

Table 4. Maximum WAR and UAR of different wavelet functions.

Family	Optimum	WAR _{max} [%]	UAR _{max} [%]
BiorSplines	bior3.5	76.8	70.1
Coiflets	coif1, coif5	77.0 (coif5)	68.9 (coif1)
Daubechies	db3, db10	78.2 (db10)	71.1 (db3)
Dmeyer	dmey	74.9	68.2
Haar	haar	66.6	61.0
ReverseBior	rbio1.5	76.4	69.7
Symlets	sym3	76.1	71.2

3.2. Wavelet Function Selection

As discussed in Section 2.2, the WPT decomposition process is based on $\Phi_{j,k}(t)$ and $\Psi_{j,k}(t)$, which are called wavelet functions. Different wavelet functions generate different descriptions for the analysed signal within the decomposition process. We explore 7 classes of available wavelet functions (totally, 54 including family members)¹ in the WPT feature extraction. The results are shown in Table 4. We observed that the ‘db10’ and ‘sym3’ function could achieve the best WAR and UAR respectively for our sample set.

3.3. Comparison with Other Features

We compare the performance of our wavelet feature set with popular further feature sets frequently used in SnS data analysis. We extract the formants (F1, F2, F3, and the corresponding amplitude energy), MFCC (Mel-frequency cepstral coefficients, 1–13), power ratio (PR800), energy ratio between the lower and higher spectrum band at 800 Hz), crest factor, and fundamental frequency (F0) from SnS and calculate the statistical values with the same approach described in Section 2.2. For the detailed basic frame-based feature extraction algorithm the reader is referred to [24], and [25]. Figure 2 shows that the proposed wavelet based feature set outperforms other frequently-used features in our set of samples. Formants could reach a relatively high WAR of 73.0%; however, other widely used feature sets can only achieve the WARs below 65% and the UARs are all below 60%. Ng et al. indicated that formants could carry important information on the upper airway structure variations [26]; therefore, further studying the relationship between formant properties and the physiological changes in the upper airway appears significant. F0, MFCC, and power ratio features are all based on Fourier Transformation (FT) [27]. Compared to FT, the Wavelet Transformation is more suitable to analyse non-stationary signals [17], due to its good frequency resolution along with finite time resolution [15].

Note that, at this stage of our work, we put less emphasis on the computational costs in our feature extraction phase. We can see that, among the feature sets, our proposed wavelet feature set contains a large feature dimension (4185), considerably higher than others, namely, formants (162), MFCC (351), power ratio (27), crest factor (27) and fundamental frequency (27). The classification performance is our major variable of interest in this work. Table 4 and Figure 2 demonstrate that, in our experiments, the average recognition performance of wavelet feature sets exceeds those of the other sets of features.

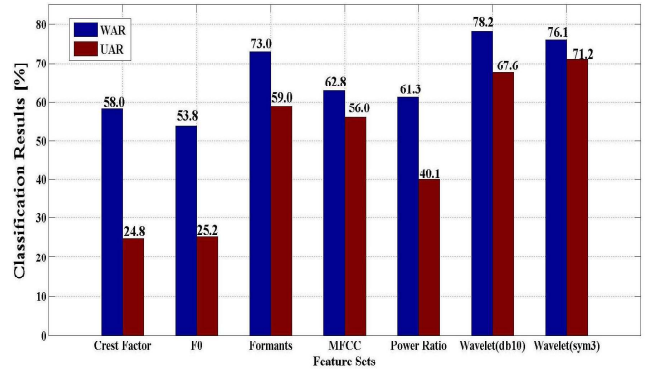


Fig. 2. VOTE SnS classification WAR and UAR by different feature sets in 2-fold cross validation.

4. CONCLUSIONS

In this study, we proposed a method based on WPT to extract wavelet features from SnS data to classify VOTE SnS with a data-trained SVM classifier. The labelled SnS data from 24 snoring subjects were grouped into two subject-independent partitions. We trained and tested our proposed classifier model with a 2-fold cross validation. Results showed that, with a suitable wavelet function, we can achieve a WAR of 78.2%, and a UAR of 71.2%, which outperformed other frequently-used features such as formants, MFCC, power ratio, crest factor, and fundamental frequency. A limitation of our study is that it is based on a relatively small cohort of subjects. Even though the number of investigated acoustic events (snoring episodes) is sufficient, the number of actual subjects is low. This bears the risk of assessing the individual acoustic characteristics of snoring subjects, rather than revealing features that are generally representative for snore sounds of the different classes. Using data from more patients can overcome this limitation. The performance of our classifier for this four-class learning problem still can be further improved. We observed formants to be comparatively relevant for classification of SnS while having a considerably lower dimension (162) if compared to wavelet features (4185). Therefore, a fusion and feature reduction phase of wavelet features and formants could be suitable to train a more robust classifier. Further, we could improve the baseline by involving balancing and instance sub-sampling training strategies.

5. RELATION TO PRIOR WORK

The WPT based feature extraction method is based on the algorithm presented by Khushaba et al. [15], which was originally used for classification of EEG, EOG, and ECG signals. We adopt the framework of our efficient toolkit openSMILE [20] to enlarge the wavelet feature set. This work is a further step in continuation of part of the authors’ previous multi-feature analysis of OSA SnS [28] and the SnS classification work in [29] and [25].

6. ACKNOWLEDGEMENTS

This work is supported by China Scholarship Council (CSC), and the European Unions’s Seventh Framework and Horizon 2020 Programmes under grant agreements No. 338164 (ERC Starting Grant iHEARu) and No. 645378 (ARIA-VALUSPA).

7. 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. 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.
- [3] P. E. Peppard, T. Young, M. Palta, and J. Skatrud, "Prospective study of the association between sleep-disordered breathing and hypertension," *New England Journal of Medicine*, vol. 342, no. 19, pp. 1378–1384, 2000.
- [4] J. M. Marin, S. J. Carrizo, E. Vicente, and A. G. Agusti, "Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study," *The Lancet*, vol. 365, no. 9464, pp. 1046–1053, 2005.
- [5] H. K. Yaggi, J. Concato, W. N. Kernan, J. H. Lichtman, L. M. Brass, and V. Mohsenin, "Obstructive sleep apnea as a risk factor for stroke and death," *New England Journal of Medicine*, vol. 353, no. 19, pp. 2034–2041, 2005.
- [6] P. J. Strollo Jr and R. M. Rogers, "Obstructive sleep apnea," *New England Journal of Medicine*, vol. 334, no. 2, pp. 99–104, 1996.
- [7] M. S. Aldrich, *Sleep medicine*. Transaction Publishers, 1999.
- [8] D. Pevernagie, R. M. Aarts, and M. De Meyer, "The acoustics of snoring," *Sleep Medicine Reviews*, vol. 14, no. 2, pp. 131–144, 2010.
- [9] 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 Otolaryngology & Laryngology*, vol. 112, no. 1, pp. 40–44, 2003.
- [10] 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.
- [11] P. Hill, B. Lee, J. Osborne, and E. Osman, "Palatal snoring identified by acoustic crest factor analysis," *Physiological Measurement*, vol. 20, no. 2, p. 167, 1999.
- [12] 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.
- [13] R. J. Beeton, I. Wells, P. Ebdon, 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, p. 1225, 2007.
- [14] 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.
- [15] 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.
- [16] D. Matsiki, X. Deligianni, E. Vlachogianni-Daskalopoulou, and L. J. Hadjileontiadis, "Wavelet-based analysis of nocturnal snoring in apneic patients undergoing polysomnography," in *Proceeding of IEEE 29th Annual International Conference of Engineering in Medicine and Biology Society, 2007*. IEEE, 2007, pp. 1912–1915.
- [17] R. R. Coifman, Y. Meyer, and V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and Their Applications*, 1992.
- [18] 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.
- [19] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1999.
- [20] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1459–1462.
- [21] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [22] 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>.
- [23] H. Alshaer, A. Pandya, T. D. Bradley, and F. Rudzicz, "Subject independent identification of breath sounds components using multiple classifiers," in *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*. IEEE, 2014, pp. 3577–3581.
- [24] Y. Wu, Z. Zhao, K. Qian, Z. Xu, and H. Xu, "Analysis of long duration snore related signals based on formant features," in *Proceedings of 2013 International Conference on Information Technology and Applications*. IEEE, 2013, pp. 91–95.
- [25] 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.
- [26] 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.
- [27] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Prentice Hall, 1978.
- [28] C. Janott, W. Pirsig, and C. Heiser, "Akustische analyse von schnarchgeräuschen," *Somnologie-Schlafforschung und Schlafmedizin*, vol. 18, no. 2, pp. 87–95, 2014.
- [29] K. Qian, Z. Xu, H. Xu, and B. P. Ng, "Automatic detection of inspiration related snoring signals from original audio recording," in *Proceedings of 2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP 2014)*. IEEE, 2014, pp. 95–99.