Archives of Acoustics

Vol. **48**, No. 1, pp. 3–12 (2023), doi: 10.24425/aoa.2022.142906

Research Paper

Sleep Snoring Sound Recognition Based on Wavelet Packet Transform

Li DING[1], Jianxin PENG[1]*, Xiaowen ZHANG[2], Lijuan SONG[2]

[1] *School of Physics and Optoelectronics, South China University of Technology*
Guangzhou, China

[2] *State Key Laboratory of Respiratory Disease, Department of Otolaryngology-Head and Neck Surgery*
Laboratory of ENT-HNS Disease, First Affiliated Hospital, Guangzhou Medical University
Guangzhou, China

*Corresponding Author e-mail: phjxpeng@163.com

(*received February 15, 2022; accepted August 31, 2022*)

Snoring is a typical and intuitive symptom of the obstructive sleep apnea hypopnea syndrome (OSAHS), which is a kind of sleep-related respiratory disorder having adverse effects on people's lives. Detecting snoring sounds from the whole night recorded sounds is the first but the most important step for the snoring analysis of OSAHS. An automatic snoring detection system based on the wavelet packet transform (WPT) with an eXtreme Gradient Boosting (XGBoost) classifier is proposed in the paper, which recognizes snoring sounds from the enhanced episodes by the generalization subspace noise reduction algorithm. The feature selection technology based on correlation analysis is applied to select the most discriminative WPT features. The selected features yield a high sensitivity of 97.27% and a precision of 96.48% on the test set. The recognition performance demonstrates that WPT is effective in the analysis of snoring and non-snoring sounds, and the difference is exhibited much more comprehensively by sub-bands with smaller frequency ranges. The distribution of snoring sound is mainly on the middle and low frequency parts, there is also evident difference between snoring and non-snoring sounds on the high frequency part.

Keywords: snoring recognition; wavelet packet transform; feature selection; machine learning.

1. Introduction

The obstructive sleep apnea hypopnea syndrome (OSAHS) is a chronic sleep-related disease affecting the general adult population ranging from 6% to 17% (Senaratna *et al.*, 2017), which is characterized by intermittently partial or complete collapse of the upper airway, resulting in frequently sleep-disordered breathing events (SDB). This kind of disease greatly affects the quality of life and even is an independent risk factor for diseases such as neurocognitive dysfunction, arterial hypertension, metabolic disorders, and cerebrovascular disease (Wang *et al.*, 2017; Hui *et al.*, 2015; Dafna *et al.*, 2013). The traditional and golden standard for clinically diagnosing OSAHS is Polysomnography (PSG) (Jiang *et al.*, 2020) with multiple sensors that must be directly connected to the body to monitor serious biological signals during sleep. However, the complex equipment, professional technicist, time-consuming process, and expensive cost limiting its wide application, makes OSA a significant but underestimated threat to public health (Ayas, 2013). An inexpensive and reliable technology to diagnose OSAHS is urgently needed. Studies have indicated that snoring is a typical and intuitive symptom of OSAHS (Dafna *et al.*, 2013; Hui *et al.*, 2015; Jiang *et al.*, 2020; Ng *et al.*, 2008; Senaratna *et al.*, 2017; Wang *et al.*, 2017) reported in more than 80% of OSAHS patients (Kapur *et al.*, 2002; Young *et al.*, 1997), which has been reported to be a potential method to monitor OSAHS. It is a kind of sleep-related noise caused by oscillations of the soft tissue structures in the upper airways (Lechner *et al.*, 2019) because of a reduction of the muscle tone and slackening of soft tissue narrowing down the upper airways.

Automatic extracting snoring episodes from recorded sleep sounds throughout the night, including breathing, speaking, and other noises, is the first but the most important step during the whole process of analyzing snoring sounds, which has been studied by many studies (Dafna *et al.*, 2013; Hwang *et al.*, 2015; Lim *et al.*, 2019; Ng *et al.*, 2008; Nonaka *et al.*, 2016; Wang *et al.*, 2017). Most of their studies focused on differencing snoring sounds and non-snoring sounds from acoustic features derived from the time domain, frequency domain, and time-frequency domain. Specifically, Jiang *et al.* (2020) designed a snoring sound detection system based on a non-contact microphone that extracted 127-dimensional time and spectral features to obtain an accuracy of 98.2% in the validation group. The work of Ng *et al.* (2008) achieved a high accuracy in the classification of snoring and non-snoring sounds using formant frequencies. Cavusoglu *et al.* (2007) implemented the snoring and non-snoring classification by the sub-band energy of sound episodes and robust linear regression with an accuracy of 90.2% from the combined dataset of 18 simple snorers and 12 OSAHS patients. Nonaka *et al.* (2016) developed the human auditory image model to extract snoring sounds automatically. These studies mainly focused on the low frequency part or some specific frequency bands rather than analysis of all sub-band, which might ignore the information on the high frequency part.

The wavelet transform (WT) is another way to divide the signal into sub-bands with different frequency ranges, which has been demonstrated to be effective in the speech signal and electroencephalogram processing (Li, Zhou, 2016; Wu *et al.*, 2008; Wang *et al.*, 2020). Wang *et al.* (2020) proposed a novel method of speaker-independent emotion recognition based on the wavelet packet analysis, which performed better than frequency features. Li and Zhou (2016) implemented the classification of electrocardiograms using the wavelet packet entropy and random forests. Wu *et al.* (2008) extracted features of electroencephalogram signals such as the energy of special sub-bands and corresponding coefficients of the wavelet packet decomposition, which had maximal separability according to the Fisher distance criterion. Qian *et al.* (2016; 2017) adopted energy features derived from the wavelet packet transform (WPT) to discriminate snoring sounds from different snoring sites with much better performance than features derived from time and frequency domains. These works indicate that wavelet transform works effectively in the analysis of biological signals.

To explore the relationship between snoring and non-snoring sounds on different frequency bands, an automatic snoring detection system based on WPT features with an XGBoost (Chen, Guestri, 2016) classifier was proposed in this study. The system includes three major steps. Firstly, the recorded sleep-related sounds were enhanced and segmented by a generalization subspace noise reduction algorithm, and signal presence probability based on energy, respectively. Then, WPT features from different wavelet functions and decomposition layers were extracted from segmented sound episodes and selected based on a series of correlation analyses. Finally, snoring sounds were detected from the trained XGBoost classifier. The contribution of the work incorporates: 1) it used WPT to extract sub-band features and yielded comparable accuracy in recognizing snoring sounds compared with existing related studies (Adesuyi *et al.*, 2022; Arsenali *et al.*, 2018; Jiang *et al.*, 2020; Sun *et al.*, 2022); 2) it discovered that signal would be exhibited much more comprehensively by sub-bands with smaller frequency ranges. And the difference between snoring and non-snoring sounds is getting more evident with the frequency range getting smaller, which is more beneficial for classifying; 3) it demonstrated that although the distribution of snoring sound is mainly on the low frequency part, the information on the high frequency part also cannot be ignored, where there is also evident difference between snoring and non-snoring sounds.

2. Material and methods

2.1. Data acquired

In this study, 24 subjects composed of simple snorers and OSAHS patients were selected from the First Affiliated Hospital of Guangzhou Medical University. All subjects have been informed and agreed with the monitoring process during the whole night. The detailed information about these subjects was described in Table 1. During sleeping, a microphone (RODE, NTG-3, Sydney, Australia) and a digital audio recorder (Rowland, Edirol R-44, Japan) were placed approximately 45 cm above the patient's mouth and nose to record the original sleep sound signals for approximately seven hours, with a sampling rate of 44.1 kHz and 16-bit resolution. PSG equipment (Alice-5, Pittsburgh, Pennsylvania, USA) was simultaneously used to monitor the subject's PSG signals.

Table 1. Information (gender, age, apnea/hypopnea (AHI), and body mass index (BMI)) of the subject.

	Simple	Mild	Moderate	Severe
Gender (M/F)	4/0	3/2	5/3	5/2
Age (years)	25 ± 5	35 ± 5.05	46.6 ± 11.58	49.9 ± 9.36
AHI	3.8 ± 0.73	10.8 ± 4.13	21.69 ± 3.46	36.77 ± 3.83
BMI	27.5 ± 4.57	30.24 ± 0.87	35.09 ± 1.05	39 ± 1.48

2.2. Pre-processing

During the process of recording sleep-related sounds, the noncontact nature of signal acquisition is often susceptible to external noise distortion. The additive background noise is inevitably superposed to a snoring sound, which will affect the fidelity of signals. Most studies (Dafna *et al.*, 2013; Jiang *et al.*, 2020; Lim *et al.*, 2019; Wang *et al.*, 2017) conducted the noise reduction process before its further analysis effectively. In the work of Karunajeewa *et al.* (2008), different enhancement algorithms were implemented to yield different snoring recognition results. Different from common noise suppression methods, generalized subspace snoring signal enhancement based on the noise covariance matrix estimation was implemented (Ding *et al.*, 2021). It was verified by our previous work that this algorithm could well update noise in real-time by recursive averaging its past values adjusted by a time-varying smoothing parameter controlled by the snoring signal presence probability during the noise suppression process. Objective quality measurements and the spectrum analysis demonstrated that this method could reduce most background noise with less signal distortion. Moreover, the enhanced snoring signal was detected and segmented by the signal presence probability, which is determined by the ratio between the local energy of the noisy signal and its minimum within a specified time window to detect the sound episode.

Figure 1 shows the process of pre-processing including noise reduction and episode segmentation. The segmented episodes were further labeled as snoring sounds and non-snoring sounds based on PSG signals by ear-nose-throat (ENT) experts. 26561 labeled sound episodes including 17704 snoring sounds and 8857 non-snoring sounds were obtained from all 24 subjects. All labeled sounds were randomly divided into the training and test sets with proportions of 70% and 30%, respectively.

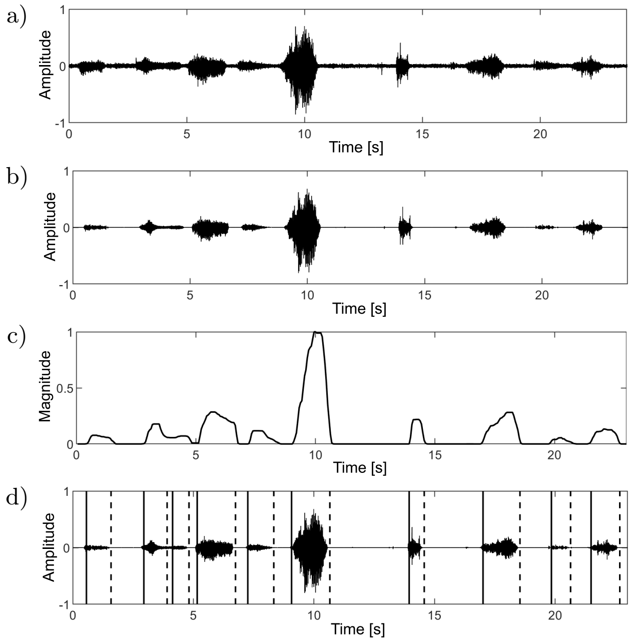

Fig. 1. Example of sound enhancement and detection: a) the original recorded noisy sound; b) the enhanced sound by subspace noise reduction algorithm; c) the signal presence probability of enhanced recorded signal; d) the detection result of sound episodes.

2.3. Feature extraction

The wavelet packet decomposition was applied to divide the signal into sub-bands with different frequency bands. Acoustic features including wavelet packet coefficients, log energy, Shannon entropy, wavelet transform cepstral coefficient, and sound pressure level based on sub-band signals were extracted for further analysis. All signals were framed by the hamming window function with 20 ms frame length and 50% overlap. Amplitude normalization was conducted to eliminate the influence of sound intensities. The statistic functions including the mean and variance of all frames in each signal were calculated to represent each signal. Table 2 shows the detailed information on the features.

Table 2. Information of the extracted features.

Feature	Description	Dimension (layer 4/layer 5/ bark sub-band)
Coefficient	Mean, variance value	32/64/34
Log-energy	Mean, variance value	32/64/34
Shannon entropy	Mean, variance value	32/64/34
Sound pressure level	Mean, variance value	32/64/34
Wavelet transform cepstral coefficient	Mean, variance value	26/26/26
Mel-frequency cepstral coefficient	Mean, variance value	26/26/26

2.3.1. Wavelet packet model

The WT (Sharma *et al.*, 2020) is a typical method to transform the time-domain audio signal into a time-frequency domain consisting of the continuous wavelet transform and discrete wavelet transform. The WT of the signal x at the time y and scale z is defined by the inner product with a wavelet function:

$$W_f(y,z) = \langle x, u_{y,z} \rangle = \frac{1}{\sqrt{y}} \int_{-\infty}^{\infty} f(t) u^* \left(\frac{(t-z)}{y} \right) \mathrm{d}t, \quad (1)$$

where $u(t)^*$ represents the complex conjugate of the wavelet function $u(t)$. WPT applies the transform step on all frequency bands. It is calculated through time-domain filtering with a sub-signal representation obtained from frequency components with each sub-band. Figure 2 shows an integrated wavelet packet tree of a signal. The original signal is first decomposed into two sub-bands in the first decomposition layer: the low frequency part-1 (1, 0) and the high frequency part-2 (1, 1). Then the low frequency part-1 and part-2 will be further decomposed with increasing decomposition layers to obtain sub-bands with much finer frequency bandwidth. The frequency bandwidth of the k-th sub-band in the j-th decomposition layer is $\frac{fs}{2^{j+1}}$ Hz; fs is the sampling rate with a value of 44.1 kHz in this work. With the increase of the decomposition layer, the finer the frequency is decomposed with much more sub-bands. There are $N_j = 2^j$ sub-bands in the j-th layer. There are 16 sub-bands in layer 4 with a bandwidth of 1378 Hz and 32 sub-bands in layer 5 with a bandwidth of 689 Hz. Moreover, we constructed bark sub-bands from decomposition layers 4 and 5 of WPT with a dimension of 17, which is based on auditory characteristics of humans. The detailed composition of bark sub-bands is shown in the last layer of Fig. 2.

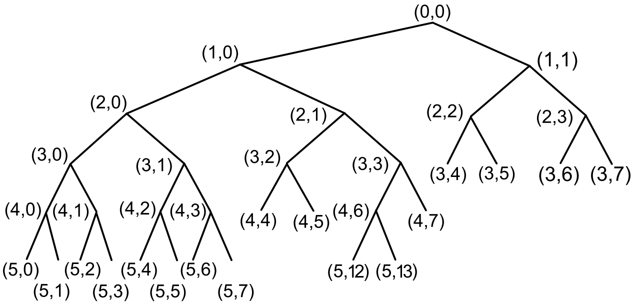

Fig. 2. Decomposition process of WPT and the construction of wavelet bark.

In the work of Monson *et al.* (2014) "high frequency" referred to a frequency above about 5 kHz which has traditionally been neglected in speech research. The "middle and low frequency" is defined below 5 kHz in this paper. As Fig. 2 shows, the bark sub-band is constructed by sub-bands in layer 4 and layer 5 which is considered a common division method for audio signal processing that accurately matches the human ear's auditory perception chrematistics (Korniienko, Machusky, 2018). The bark and layer 5 decomposition structure have the same sub-band distributions in the middle and low frequency part (0–5.5 kHz) composed by (5,0)–(5,7), while layer 5 has much finer sub-bands in the high frequency. Studies have indicated that the energy of snoring sounds is mainly concentrated below 2 kHz (Perez-Padilla *et al.*, 1993). The frequency range of the first sub-band at layer 4 is 0–1378 Hz that most snoring information locate in this frequency band. To explore the influence of the frequency bandwidth of the sub-band on the classification result, and the difference between snoring and non-snoring sounds in the middle and low frequency part and high frequency part, acoustic features extracted from layer 4, layer 5, and bark sub-bands are discussed in the work.

2.3.2. Wavelet packet coefficients

The coefficients by WPT can reveal the local characteristics of signals. The mean values of the coefficients of the k-th sub-band in the j-th layer are described as:

$$w_{j,k} = \frac{\sum\limits_{n} v_{j,k,n}}{N_k}, \quad n = 1,2,...,N_k, \;\; k = 1,2,...,2^j, \tag{2}$$

where N_k is the number of the coefficient component in the k-th sub-band with the value of 882; $v_{j,k,n}$ represents the $n-$th coefficient component of the k-th sub-band in the j-th decomposition layer. There are $N_j = 2^j$ coefficients in the j-th decomposition, which are $w_{j,1}, w_{j,2}, ..., w_{j,2^j}$.

2.3.3. Log-energy

The log-energy of the k-th sub-band signal in the j-th level can be calculated by:

$$\log E_{j,k} = \sum_{n=1}^{N_k} v_{j,k,n}^2. \tag{3}$$

2.3.4. Shannon entropy

The probability of the n-th coefficient at its corresponding node can be calculated by:

$$p_{j,k,n} = \frac{E_{j,k,n}}{E_{j,k}} = \frac{v_{j,k,n}^2}{\sum\limits_{n=1}^{N_k} v_{j,k,n}^2}, \tag{4}$$

moreover, Shannon entropy (SE) is defined by the probability distribution of energy $p_{j,k,n}$ as Eq. (4), which is a measure of uncertainty associated with random variables in information theory:

$$\text{SE}_{j,k} = -\sum_{n=1}^{N_k} p_{j,k,n} \cdot \log\left(p_{j,k,n}\right). \tag{5}$$

2.3.5. Wavelet transform cepstral coefficient

WPT can be treated as a filter to divide the frequency to some sub-bands with the equal bandwidth, just like Mel-filter. Then the cepstral coefficient of the signal after the WPT filter can be calculated using discrete cosine transform (DCT), which is called the wavelet transform cepstral coefficient (WTCC):

$$\text{WTCC}_m = \sqrt{\frac{2}{N_j}} \sum_{j=0}^{N_j} \log\left(E_{j,k}\right) \cos\left(\frac{\pi m\,(2j-1)}{2N_j}\right), \tag{6}$$

where N_j is the number of sub-bands in the j-th layer; m indicates the m-th DCT spectral line, which was set as 13 in the work.

2.3.6. Mel-frequency cepstral coefficients

The study (Jiang *et al.*, 2020) has indicated that there are obvious differences between snoring and non-snoring sounds via Mel-frequency cepstral coefficients (MFCC). The MFCC of the original signal with 13-dimension is also extracted for snoring sound recognition.

2.4. Wavelet function

As described in Eq. (1), the WPT is based on the wavelet function. Different wavelet functions may result in different WPT features. There are many wavelet function families such as BiorSplines, Coiflets, Daubechies, Symlets, and so on. The Daubechies wavelet family (DB1, DB2, DB3, DB4, DB5, DB6, DB7, DB8, DB9, DB10) has been widely used in the processing of speech and other biological signals. Also, the work of Qian *et al.* (2017) indicated that the Daubechies function (DB3, DB10) performed better on recognition of a snoring site. In this paper, we explore their performance in differentiating snoring and non-snoring sounds.

3. Classification and result

3.1. Classification model

In the study, XGBoost classifier was adopted in this study. XGBoost is an improved algorithm with good performance and high efficiency based on the gradient boosting decision that can construct boosted trees efficiently and operate in parallel. The core of the algorithm is the optimization of the value of the objective function (Torlay *et al.*, 2017). The parameter of the XGBoost classifier is essential for classification performance. Based on a 10-fold cross-validation of the training set, the optimal parameter was obtained. The number of base trees was set as 400, the max depth of trees was 6, and the learning rate was 0.3. Other parameters were set as the default value of XGBoost in Scikit-learn (Pedregosa *et al.*, 2011).

3.2. Feature selection

Feature selection is a vitally important step during the classification task because it can reduce the redundancy of features to improve the robustness of the model and reduce the computation complexity. In this paper, feature selection based on the correlation analysis is conducted to select distinguishing features. Two Pearson correlation coefficients were calculated including correlation coefficients between features and their related labels with a value of P1 and correlation coefficients among features with a value of P2. Features with high correlation with labels and low correlation with other features were selected by thresholds a and b, respectively. There were two steps for feature selection. Firstly, features were reserved if P1 was higher than a. Then, the reserved features were dropped out if P2 was higher than b to obtain relatively independent features. To fully make use of the limited dataset, the 10-fold cross-validation was used in the training set to optimize the model and select features. The threshold a and b were obtained by experiment to set as 0.8 and 0.7, respectively. Moreover, the effect of the decomposition levels and wavelet functions on the classification of snoring and non-snoring sound is explored.

3.3. Model evaluation

To evaluate the performance of the proposed recognition system of snoring sound, evaluating indexes such as sensitivity, accuracy, precision, and F1 score are expressed as follows:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})}, \tag{7}$$

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}, \tag{8}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}, \tag{9}$$

$$\text{F1} = \frac{2\text{Precision} \cdot \text{Sensitivity}}{(\text{Precision} + \text{Sensitivity})}, \tag{10}$$

where TP represents the number of snoring sounds classified as snoring sounds (true positive), TN is the number of non-snoring sounds truly detected as non-snoring sounds (true negative), FP represents the number of non-snoring sounds falsely recognized as snoring sounds (false positive), and FN is the number of events corresponding to the false detection of snoring sound as non-snoring sound (false negative).

3.4. Classification results

Figure 3 shows the distribution of coefficient 1 and WTCC13 which have the first and second highest coefficients with labels selected by correlation analysis. 5000 samples were randomly selected from 24 subjects in the training set. It shows that the distribution of coefficient 1 and WTCC13 of snoring sounds is different from non-snoring sounds, which could distinguish snoring sounds to a certain extent. Figure 4 shows the overall accuracy of snoring and non-snoring sounds with different wavelet functions under different decomposition layers. The WPT features extracted from different decomposition layers and different Daubechies wavelet functions could work well with accuracies more than 94%.

It can be seen from Fig. 4a that the overall accuracies of WPT features extracted from layer 5 are slightly 0.5 percentage points higher than accuracies

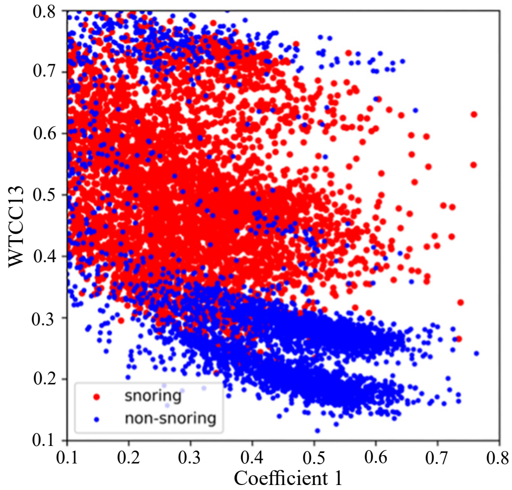

Fig. 3. The distribution of coefficient 1 and WTCC13 of decomposition layer 5 randomly selected from the training set of 5000 samples. The coefficient 1 and WTCC13 are the coefficient and WTCC of the first sub-band and thirteenth sub-band components. Snoring and non-snoring segments are denoted by red circle and blue circle symbols, respectively.

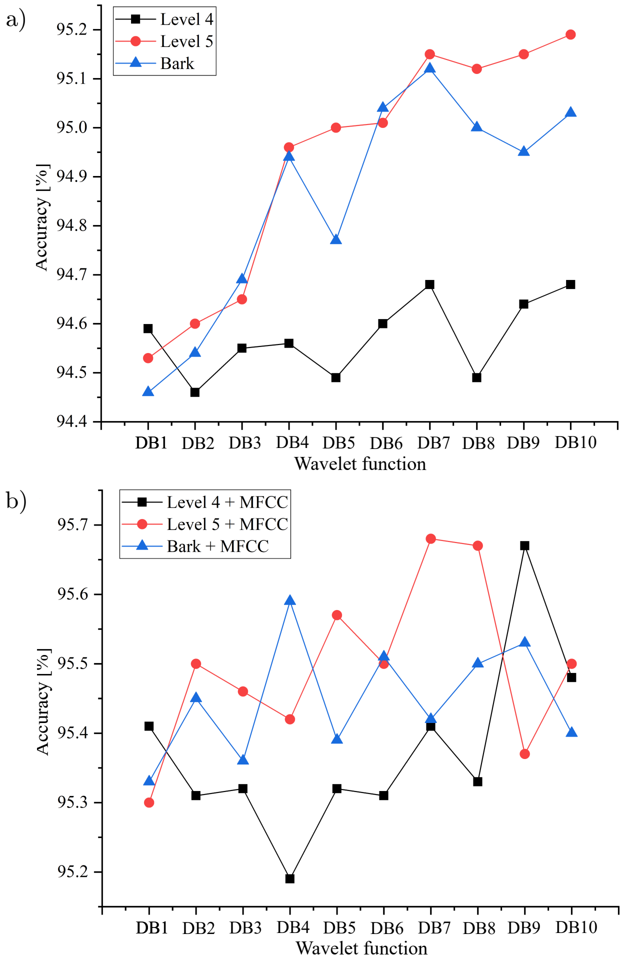

Fig. 4. The snoring recognition result of two kinds of feature sets under decomposition layer 4, 5, and bark and wavelet function DB1 to DB10 in Daubechies family: a) WPT feature set; b) MFCC and WPT combined feature sets.

of layer 4 under most wavelet functions. With the increase of the decomposition level, the frequency of the signal is divided into smaller bandwidths to obtain much more detailed information about signals, which is beneficial for distinguishing snoring and non-snoring sounds, and yield much better performance. Comparing the recognition rate of layer 5 and bark sub-bands, there is little difference between these two decomposition structures. However, layer 5 and the bark sub-band have the same decomposition construction on low frequencies (0–5.5 kHz), and layer 5 has much finer division than the bark sub-band on high frequencies (5.5–44.1 kHz). In other words, the bark sub-band puts much more emphasis on low frequency. The result shows that the energy of snoring sound and the difference between snoring and non-snoring are mainly on low frequencies. The information on the high frequencies part also cannot be ignored.

It also can be observed from Fig. 4a that the wavelet function also influences the final classification result. The overall accuracies are different between wavelet functions in the same decomposition layer. DB7, DB9, and DB10 in level 5, and DB7 in bark yield much higher recognition for WPT features among all test Daubechies wavelet functions, which are 95.2% approximately, indicating that the wavelet function plays an important role in the decomposition of the signal. Comparing Fig. 4, the WPT and MFCC combined features yielded accuracy with an average value of 95.5%, much better than simple WPT features in terms of overall accuracies under all test wavelet functions and decomposition levels. The difference in recognition results of the MFCC and WPT combined features between decomposition levels and wavelet functions is not as obvious as simple WPT features. The first three recognition rates are DB7 in level 5, DB8 in level 5, and DB9 in L4 which are 95.68%, 95.67%, and 95.65% respectively under all test conditions.

Table 3 shows detailed results for snoring and non-snoring recognition of different feature sets including MFCC, WTCC, WPT features, and WPT + MFCC combined features under the selected DB7, DB8, DB9, and DB10 wavelet functions in decomposition layer 5 and bark sub-bands. It can be known from Table 3 that for WPT features, the accuracy of the three kinds of selected feature sets is comparable, which is around 95.2%. However, the features extracted from DB7-Level 5 yield much higher sensitivity with 96.94%, indicating a higher probability of real snoring sound being recalled. For WPT – MFCC combined feature sets, the features extracted from level 5 with the wavelet function DB7 achieved the best performance during all test conditions considering sensitivity, precision, and F1-score, which are 97.27%, 96.48%, and 96.88%, respectively. Compared with MFCC, simple WPT features, the WPT – MFCC features performed best, and there is an average improvement of 1 percent-

Table 3. Detailed classification results and dimensions of selected features of different feature sets including Cepstral coefficients features, WPT features, and WPT + MFCC combined features under the selected wavelet functions and decomposition layers selected from Fig. 4.

Feature-set	Number of features (total)	Accuracy [%]	Sensitivity [%]	Precision [%]	F1 [%]	AUC
Cepstral coefficients						
MFCC	26	94.76	96.58	95.83	96.20	0.98
WTCC	26	93.57	96.17	94.58	95.37	0.98
WPT						
DB7-L5	69 (256)	95.15	96.94	96.21	96.50	0.99
DB9-L5	82 (256)	95.15	96.81	96.18	96.49	0.99
DB10-L5	81 (256)	95.20	96.81	96.23	96.52	0.99
WPT + MFCC						
DB7-L5	98 (282)	95.68	97.27	96.48	96.88	0.99
DB8-L5	96 (282)	95.67	97.44	96.30	96.87	0.99
DB9-L4	99 (162)	95.65	97.27	96.27	96.77	0.99

age points and 0.5 percentage points for WPT – MFCC combined features in terms of overall accuracy, recall, precision, and F1-score mentioned in the work. The dimensions of selected features based on the 10-fold cross-validation are also displayed in Table 3. Features with little contribution to the classification are dropped by the selection technology based on correlation coefficients. Based on the aforementioned discussion, the features extracted from the wavelet function DB7 in level 5 decomposition performed better considering simple WPT features and WPT – MFCC combined features with dimensions of 69 and 98, respectively.

We also compared two kinds of cepstral coefficients MFCC and WTCC, which are derived from the Mel-frequency filter and the wavelet packet transform filter respectively. The MFCC outperformed the WTCC with an average improvement of 1 percentage points in terms of evaluation standards mentioned in the work, which means that the Mel-frequency could carry more important information on the upper airway structure variations than wavelet packet transform does.

4. Discussion

In this work, we proposed a novel system to automatically extract snoring sounds from the recorded sounds during sleep based on WPT features. Based on wavelet packet transform, the snoring sound was decomposed into different sub-bands with the same bandwidth and different frequency ranges. Results of WPT features indicated that the information on snoring sounds and the difference between snoring and non-snoring sounds were mainly in the middle and low frequency. With increasing decomposition layer, the signal was decomposed with much smaller sub-bands, and the difference between snoring and non-snoring sounds was much more obviously accompanied by a higher classification accuracy. The snoring sound detection is the first but vital step during the whole analysis system of snoring sounds. Many studies have detected snoring episodes from different kinds of domains.

In previous studies (Han *et al.*, 2006; Karunajeewa *et a*l., 2011; Qian *et al.*, 2015; Solà-Soler *et al.*, 2007; Sun *et al.*, 2022), acoustic features extracted from frequency sub-band of the signal have been demonstrated effectively and widely used in classifying snoring and non-snoring sounds. Qian *et al.* (2015) used the 1000 Hz sub-band features, and power ratio to detect snoring sound segments. Cavusoglu *et al.* (2007) explored the sub-band energy distribution of snoring and non-snoring segments by dividing the 0–7500 Hz frequency range into 500 Hz sub-bands, which yielded 90.2% accuracy for simple snorers. These works indicated that the information distribution of snoring sounds is different among sub-bands, which mainly focus on middle and low frequency parts. Moreover, since the bark sub-bands focus on the low frequency part and are sparse in the high frequency part. The WPT furtherly divided the high frequency part based on bark sub-bands which makes the difference in the high frequency much more obviously. And recognition accuracies of layer 5 are slightly higher than bark sub-bands under most wavelet function test conditions. Although the distribution of snoring is mainly concentrated in the middle and low frequency parts, the information of snoring sounds in high frequency part also cannot be ignored. And with increasing of decomposition layer, the difference between snoring and non-snoring sounds is getting obvious, because the signal will be exhibited much more comprehensively by sub-bands with smaller frequency ranges.

The results of works (Cavusoglu *et al.*, 2007; Duckitt *et al.*, 2006; Emoto *et al.*, 2018; Jiang *et al.*, 2020; Lim *et al.*, 2019; Qian *et al.*, 2015; Sun *et al.*, 2022) showed that MFCC and its related features could yield relatively good performance on recognizing snoring sounds. Jiang *et al.* (2020) extracted the Mel-spectrogram of signal and used a convolution neural network (CNN) to classify snoring and non-snoring sounds with good performance. Sun *et al.* (2022) also indicated that different components of MFCC yielded different contributions to the classification result. The same result is also shown in this study. The high effectivity of MFCC and WTCC may be caused by dividing the signal into different sub-bands with the same frequency range through Mel-filter banks based on the mechanism of human hearing.

There are many studies to classify snoring and non-snoring sounds from other parts shown in Table 4 (Adesuyi *et al.*, 2022; Ankişhan, Tuncer, 2017; Arsenali *et al.*, 2018; Jiang *et al.*, 2020; Lim *et al.*, 2019; Nonaka *et al.*, 2016; Sun *et al.*, 2022) used the processing way of images to analyze snoring and non-snoring sound with the recognition accuracy of 95.1% in Mel-spectrogram. In the work of Nonaka *et al.* (2016), the auditory image model, which has been used to numerically explain the auditory phenomenon of human's auditory, was developed to automatically extract snoring sounds from sleep sounds, which could achieve a sensitivity of 97.2% from 40 subjects. The work of Lim *et al.* (2019) and Adesuyi *et al.* (2022) yielded the highest 99.0% accuracy compared with all other studies. However, these results are not convincing because that there are only 8 and 6 subjects for the two studies. The diversity of samples too small to demonstrate the effectiveness of the proposed algorithm. All these studies are based on subject dependence, which cannot be directly used in practice. It demonstrated evident differences between snoring and non-snoring sounds from all kinds of aspects. The models based on deep learning perform much better than traditional machine learning, which also demonstrates the effectiveness of WPT features proposed in this work. It is worth noticing that these works have yielded competitive results on their own limited dataset. However, there is no sense to compare these accuracies because of the inconsistent dataset used in studies. The dataset of each study is established by its own team with subjects from different counties, different recording equipment, and different labeling standards of snoring and non-snoring sounds. But the result of our work is comparable with previous studies in terms of our own dataset. And it demonstrated that the distribution of snoring and non-snoring sounds in each sub-band of all frequency ranges is obviously different.

In conclusion, there are some contributions to this study. Firstly, it used WPT to extract sub-band features and yielded comparable accuracy in recognizing snoring sounds. Then, it discussed different wavelet functions and decomposition layers, concluding that the difference is getting more evident with the frequency range getting smaller, which is more beneficial for classifying. Thirdly, it demonstrated that although the distribution of snoring sounds is mainly on the low frequency part, there are also differences between snoring and non-snoring sounds in the high frequency part. There are some limitations of the work. The data partition methods mainly included subject dependence and subject independence which greatly influenced the classification performance. Subject dependence is an original data partition method to discuss the features' influence on classification performance, while the result of subject independence is more suitable for use in practice. In this work, the partition of the training set and test set is based on subject dependence because of the limited subjects. The subject independent classification must base on a huge number of training and validation subjects to make up for the influence of individual characteristics. There are only 24 subjects used in the study, which is hard to perform subject independent classification considering individual characteristics. It is the next step of the paper to collect much more snoring sounds from different subjects to implement detecting snoring sounds based on subject independence.

Table 4. Classification results of current studies recorded by ambient microphones. Abbreviations include AdaBoost (adaptive boosting), CNN (convolution neural network), MNLR (multi-nominal logistic regression), STFT (short-time Fourier transform), RNN (recurrent neural network), and LLEs (largest Lyapunov exponents).

Author	Subjects	Features + classifier	Accuracy [%]	Sensitivity [%]
Jiang *et al.* (2020)	15	Mel-spectrogram + CNN	95.1	95.4
Nonaka *et al.* (2016)	40	Audio image model + MNLR	97.3	97.2
Sun *et al.* (2022)	24	Sub-band features + XGBoost	94.3	96.5
Lim *et al.* (2019)	8	MFCC, STFT + RNN	98.5	99.3
Ankişhan, Tuncer (2017)	22	Chaotic features + LLEs	94.4	88.3
Arsenali *et al.* (2018)	20	MFCC + RNN	95.0	92.0
Adesuyi *et al.* (2022)	6	MFCC + CNN	99.0	
This work	24	WPT features + XGBoost	95.15	96.94
		WPT + MFCC + XGBoost	95.68	97.27

5. Conclusion

This study proposed a snoring sounds recognition system based on WPT features and XGBoost classifier. The recorded sleep sounds of 24 subjects, firstly were enhanced and segmented by a subspace noise reduction algorithm and signal presence probability based on the estimation of noise autocorrelation respectively to obtain potential snoring episodes. In the training set, 10-fold cross-validation was implemented to select appreciated features and models. Results of the recognition system showed that features based on sub-bands could well classify snoring and non-snoring sounds with accuracy of 95.65%, sensitivity of 97.27%, and precision of 96.58% in the test set for DB7 function and level 5, the best combination of all test conditions. And the comparison among decomposition layers shows, although the distribution of snoring sounds is mainly in the low frequency part, there is also evident difference between snoring and non-snoring sounds in the high frequency part. However, the MFCC – WPT combined feature set outperformed the simple MFCC and WPT feature sets, with accuracy of 95.68%, sensitivity of 97.27%, and precision of 96.68%. These results have demonstrated that the wavelet packet analysis is effective in recognizing snoring sounds with less computational complexity, which can be further developed to analyze OSAHS at home.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers 11974121, 81570904) and National Youth Foundation of China (grant number 81900927).

Ethical approval

This work was approved by the Ethics Committee of Guangzhou Medical University, and an informed consent was obtained from each participant. This work does not contain any studies with animals performed by any of the authors.

References

1. Adesuyi T.A., Kim B.M., Kim J. (2022), Snoring sound classification using 1D-CNN model based on multi-feature extraction, *International Journal of Fuzzy Logic and Intelligent Systems*, **22**(1): 1–10, doi: 10.5391/IJFIS.2022.22.1.1.
2. Ankişhan H., Tuncer A.T. (2017), A new portable device for the snore/non-snore classification, [in:] *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, doi: 10.1109/ICEngTechnol. 2017.8308212.
3. Arsenali B. *et al.* (2018), Recurrent neural network for classification of snoring and non-snoring sound events, [in:] *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 328–331, doi: 10.1109/EMBC. 2018.8512251.
4. Ayas N.T. (2013), *Risk factors for obstructive sleep apnea*, [in:] Encyclopedia of Sleep, pp. 212–214, doi: 10.1016/B978-0-12-378610-4.00308-9.
5. Cavusoglu M., Kamasak M., Erogul O., Ciloglu T., Serinagaoglu Y., Akcam T. (2007), An efficient method for snore/nonsnore classification of sleep sounds, *Physiological Measurement*, **28**(8): 841–853, doi: 10.1088/ 0967-3334/28/8/007.
6. Chen T., Guestrin C. (2016), XGBoost: A scalable tree boosting system, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, doi: 10.1145/ 2939672.293 9785.
7. Dafna E., Tarasiuk A., Zigel Y. (2013), Automatic detection of whole night snoring events using noncontact microphone, *PLOS ONE*, **8**(12): e84139, doi: 10.1371/journal.pone.0084139.
8. Ding L., Peng J., Jiang Y., Song L. (2021), Generalized subspace snoring signal enhancement based on noise covariance matrix estimation, *Circuits, Systems, and Signal Processing*, **40**(7): 3355–3373, doi: 10.1007/ s00034-020-01623-3.
9. Duckitt W.D., Tuomi S.K., Niesler T.R. (2006), Automatic detection, segmentation and assessment of snoring from ambient acoustic data, *Physiological Measurement*, **27**(10): 1047–1056, doi: 10.1088/0967-3334/27/10/010.
10. Emoto T., Abeyratne U.R., Kawano K., Okada T., Jinnouchi O., Kawata I. (2018), Detection of sleep breathing sound based on artificial neural network analysis, *Biomedical Signal Processing and Control*, **41**: 1–89, doi: 10.1016/j.bspc.2017.11.005.
11. Han W., Chan C.F., Choy C.S., Pun K.P. (2006), An efficient MFCC extraction method in speech recognition, [in:] *2006 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 145–148, doi: 10.1109/ iscas.2006.1692543.
12. Hui J. *et al.* (2015), Acoustic analysis of snoring in the diagnosis of obstructive sleep apnea syndrome: A call for more rigorous studies, *Journal of Clinical Sleep Medicine*, **11**(7): 765–771, doi: 10.5664/jcsm.4856.
13. Hwang S.H. *et al.* (2015), Polyvinylidene fluoride sensor-based method for unconstrained snoring detection, *Physiological Measurement*, **36**(7): 1399–1414, doi: 10.1088/0967-3334/36/7/1399.
14. Iber C., Ancoli-Israel S., Chesson A., Quan S.F. (2007), *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*, Westchester, Illinois, American Academy of Sleep Medicine.
15. Jiang Y., Peng J., Zhang X. (2020), Automatic snoring sounds detection from sleep sounds based on deep learning, *Physical and Engineering Sciences in Medicine*, **43**(2): 679–689, doi: 10.1007/s13246-020-00876-1.

16. Kapur V., Strohl K.P., Redline S., Iber C., O'Connor G., Nieto J. (2002), Underdiagnosis of sleep apnea syndrome in U.S. communities, *Sleep and Breathing*, **6**(2): 49–54, doi: 10.1007/s11325-002-0049-5.

17. Karunajeewa A.S., Abeyratne U.R., Hukins C. (2008), Silence-breathing-snore classification from snore-related sounds, *Physiological Measurement*, **29**(2): 227–243, doi: 10.1088/0967-3334/29/2/006.

18. Karunajeewa A.S., Abeyratne U.R., Hukins C. (2011), Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome, *Physiological Measurement*, **32**(1): 83–97, doi: 10.1088/0967-3334/32/1/006.

19. Korniienko O., Machusky E. (2018), Voice activity detection algorithm using spectral-correlation and wavelet-packet transformation, *Radioelectronics and Communications Systems*, **61**(5): 185–193, doi: 10.3103/S0735272718050011.

20. Lechner M., Breeze C.E., Ohayon M.M., Kotecha B. (2019), Snoring and breathing pauses during sleep: interview survey of a United Kingdom population sample reveals a significant increase in the rates of sleep apnoea and obesity over the last 20 years – data from the UK sleep survey, *Sleep Medicine*, **54**: 250–256, doi: 10.1016/j.sleep.2018.08.029.

21. Li T., Zhou M. (2016), ECG classification using wavelet packet entropy and random forests, *Entropy*, **18**(8): 1–16, doi: 10.3390/e18080285.

22. Lim S.J., Jang S.J., Lim J.Y., Ko J.H. (2019), Classification of snoring sound based on a recurrent neural network, *Expert Systems with Applications*, **123**: 237–245, doi: 10.1016/j.eswa.2019.01.020.

23. Monson B.B., Hunter E.J., Lotto A.J., Story B.H. (2014), The perceptual significance of high-frequency energy in the human voice, *Frontiers in Psychology*, **5**: 587, doi: 10.3389/fpsyg.2014.00587.

24. Ng A.K., Koh T.S., Baey E., Lee T.H., Abeyratne U.R., Puvanendran K. (2008), Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?, *Sleep Medicine*, **9**(8): 894–898, doi: 10.1016/j.sleep.2007.07.010.

25. Ng A.K., Koh T.S., Puvanendran K., Abeyratne U.R. (2008), Snore signal enhancement and activity detection via translation-invariant wavelet transform, *IEEE Transactions on Biomedical Engineering*, **55**(10): 2332– 2342, doi: 10.1109/TBME.2008.925682.

26. Nonaka R. *et al.* (2016), Automatic snore sound extraction from sleep sound recordings via auditory image modeling, *Biomedical Signal Processing and Control*, **27**: 7–14, doi: 10.1016/j.bspc.2015.12.009.

27. Pedregosa F. *et al.* (2011), Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, **12**(85): 2825–2830.

28. Perez-Padilla J.R., Slawinski E., Difrancesco L.M., Feige R.R., Remmers J.E., Whitelaw W.A. (1993), Characteristics of the snoring noise in patients with and without occlusive sleep apnea, *American Review of Respiratory Disease*, **147**(3): 635–644, doi: 10.1164/ajrccm/147.3.635.

29. Qian K. *et al.* (2017), Snore sound recognition: On wavelets and classifiers from deep nets to kernels, [in:] *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3737–3740, doi: 10.1109/EMBC.2017.8037669.

30. Qian K., Janott C., Zhang Z., Heiser C., Schuller B. (2016), Wavelet features for classification of vote snore sounds, [in:] *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 221–225, doi: 10.1109/ICASSP.2016.7471669.

31. Qian K., Xu Z., Xu H., Wu Y., Zhao Z. (2015), Automatic detection, segmentation and classification of snore related signals from overnight audio recording, *IET Signal Processing*, **9**(1): 21–29, doi: 10.1049/iet-spr.2013.0266.

32. Senaratna C.V. *et al.* (2017), Prevalence of obstructive sleep apnea in the general population: A systematic review, *Sleep Medicine Reviews*, **34**: 70–81, doi: 10.1016/j.smrv.2016.07.002.

33. Sharma G., Umapathy K., Krishnan S. (2020), Trends in audio signal feature extraction methods, *Applied Acoustics*, **158**: 107020, doi: 10.1016/j.apacoust.2019.107020.

34. Solà-Soler J., Jané R., Fiz J.A., Morera J. (2007), Automatic classification of subjects with and without Sleep Apnea through snoring analysis, [in:] *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6093–6096, doi: 10.1109/IEMBS.2007.4353739.

35. Sun X., Peng J., Zhang X., Song L. (2022), Effective feature selection based on Fisher Ratio for snoring recognition using different validation methods, *Applied Acoustics*, **185**: 108429, doi: 10.1016/j.apacoust.2021.108429.

36. Torlay L., Perrone-Bertolotti M., Thomas E., Baciu M. (2017), Machine learning–XGBoost analysis of language networks to classify patients with epilepsy, *Brain Informatics*, **4**: 159–169, doi: 10.1007/s40708-017-0065-7.

37. Wang C., Peng J., Song L., Zhang X. (2017), Automatic snoring sounds detection from sleep sounds via multi-features analysis, *Australasian Physical and Engineering Sciences in Medicine*, **40**: 127–135, doi: 10.1007/s13246-016-0507-1.

38. Wang K., Su G., Liu L., Wang S. (2020), Wavelet packet analysis for speaker-independent emotion recognition, *Neurocomputing*, **398**: 257–264, doi: 10.1016/j.neucom.2020.02.085.

39. Wu T., Yan G.-Z., Yang B.-H., Sun H. (2008), EEG feature extraction based on wavelet packet decomposition for brain computer interface, *Measurement: Journal of the International Measurement Confederation*, **41**(6): 618–625, doi: 10.1016/j.measurement.2007.07.007.

40. Young T., Evans L., Finn L., Palta M. (1997), Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women, *Sleep*, **20**(9): 705–706, doi: 10.1093/sleep/20.9.705.