A Machine-Learning-Based Detection Method for Snoring and Coughing

Chun-Hung Yang[1], Yung-Ming Kuo[2*], I-Chun Chen[3], Fan-Min Lin[4], Pau-Choo Chung[4]

[1] Dept. of Electronic Engineering, Southern Taiwan University of Science and Technology, Taiwan
[2] Dept. of Electronic Engineering, National Formosa University, Taiwan
[3] Center For General Education, National Formosa University, Taiwan
[4] Dept. of Electrical Engineering, National Cheng Kung University, Taiwan
eliyang@stust.edu.tw, ymkuo@gs.nfu.edu.tw, yjchen@gs.nfu.edu.tw, a90167130@gmail.com, pcchung@ee.ncku.edu.tw

Abstract

Poor sleep quality is a common disease for modern people. Snoring is one of the essential indicators to measure Obstructive Sleep Apnea (OSA). When sleeping, the number of episodes of snoring and coughing are related to the estimated sleep quality. This study proposes a method to detect snoring and coughing in patients when sleeping. The proposed method includes three stages. Firstly, the nightly sound data for a patient are segmented to each independent event. Secondly, the time domain signal is changed to a frequency domain signal by Fourier Transform, and then the features are extracted from the snoring and coughing episodes. Lastly, the Support Vector Machine (SVM) and the Hidden Markov Model (HMM) are used to recognize snoring and coughing. The result of our experiment demonstrates that this method has good detection performance.

Keywords: Coughing detection, Snoring detection, Machine learning, Hidden Markov Model, Support Vector Machine

1 Introduction

Poor sleep quality and sleep apnea are common diseases for modern people. As the National Institute of Health reports, sleep is a vital part of people's daily routine [1]. In addition to feeling tired quickly during the day, people with poor sleep quality are more likely to have lower immunity and suffer from physical and mental illnesses. Good sleep is strongly related to better physical, cognitive, and psychological health. In contrast, poor sleep can impair cognitive and psychological functioning and worsen general physical health [2]. For example, good sleep quality can reduce the risk of Alzheimer's disease [3].

Obstructive sleep apnea (OSA) is a common and highly prevalent disease in the general population [4-7]. An airway blockage causes OSA during sleep with repetitive apneas and hypopneas. Most OSA patients are overweight; thus, the respiratory tract becomes narrowed. Some patients are born with a small chin or hypertrophy of the tonsils [8], which makes the upper respiratory tract collapse. This would obstruct the respiratory tract and produce a shallow breathing effect. Patients with severe OSA also have a high probability of choking.

Many instruments are used to record information when a patient is sleeping. Polysomnography (PSG) test is frequently used because it contains bioelectric signals, for example, electromyography (EMG), electroencephalography (EEG), electrocardiography (ECG), and electrooculography (EOG) tests. PSGs require a large number of wired devices to be attached to the patient but it can produce accurate detection results. However, the patient's discomfort may result in a change of sleep habits and decrease sleep quality. PSGs are expensive and non-portable, so patients must visit the hospital for the examination. A lack of hospital beds in sleep centers results in waiting lists for examinations, and patients with severe OSA must wait for therapy.

Snoring is a prevalent condition caused by breathing through a narrow respiratory tract as the throat muscles relax during sleep. Almost half of men and 20%~30% of women experience snoring problems. Older people, whose throat muscles are not as elastic, can exhibit a collapsed respiratory tract during sleep. A study shows that more than 80% of men and 70% of women over 60 snore. Coughing is a natural reaction when the trachea or bronchial mucosa is stimulated due to foreign objects entering. The process of coughing includes muscle contraction. If coughing occurs during sleep, sleep is disturbed and may cause breathing abnormalities.

However, patients who suffer from OSA often experience snoring and coughing when sleeping at night. When patients first visit the hospital for treatment, doctors do not have access to any snoring or coughing data, so if such data can be accurately detected, the diagnosis will be more accurate. This study proposes a quick screening method for patients to aid in diagnosis and reduces waiting lists. This study sees patients carrying a lightweight and non-contact audio recording mechanism to record noises when sleeping. The mechanism proposed in this paper to detect snoring and coughing during sleep, as shown in Figure 1, has a recording pen with a highly sensitive microphone.

There are four stages of sound detection, as shown in Figure 2. An audio signal is segmented into independent events using endpoint detection, and different features are extracted from independent events. These features are then input into the classified model to compute the classification result.

[*] Corresponding Author: Yung-Ming Kuo; E-mail: ymkuo@gs.nfu.edu.tw

DOI: 10.53106/160792642022112306007

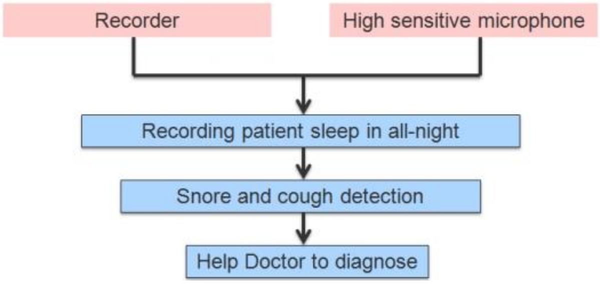

Recorder
High sensitive microphone
Recording patient sleep in all-night
Snore and cough detection
Help Doctor to diagnose

Figure 1. The mechanism for sound recording and detection during sleep

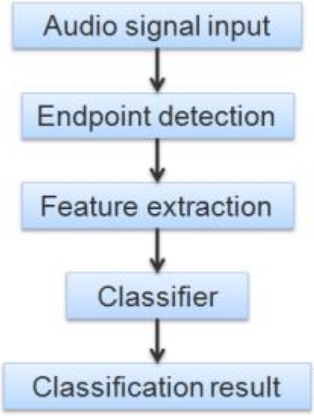


Figure 2. Sound detection process

Most endpoint detection methods use energy and a zero-crossing rate to calculate the threshold [9]. A result greater than the threshold is categorized as an audio event. Ali Azarbarzin et al. [10] proposed a Modified Vertical-Box Control Chart to replace a Vertical-Box Control Chart in statistics to segment valid audio events for subjects during sleep.

For feature extraction and classifier for snoring, W. D. Duckitt et al. [11] used the Mel-a frequency Cepstral coefficient (MFCC) as a feature with a Hidden Markov Model (HMM) for sound classification such as snoring, breathing, duvet noise, silence, and other noise. Snoring is classified with an accuracy of 82% to 89%. M. Cavusoglu et al. [9] proposed a Short-Time Fourier Transform (STFT) to calculate ten-dimensional feature vectors and used Principal Component Analysis to reduce the ten-dimensional feature vectors to two-dimensional feature vectors. A Fuzzy C-means clustering method was then used to classify snoring, breathing, and other sounds with an accuracy of 88% to 98%. A. S. Karunajeewa et al. [12] proposed four features for classifying snoring, breathing, and silence. The four features are the number of zero crossings, the energy of the signal, the normalized autocorrelation coefficient, and the first predictor coefficient for linear predictive coding with an accuracy of about 90%.

For feature extraction and a classifier for coughing, Sung-Hwan Shin et al. [13] proposed an automatic cough detection system to monitor a person’s physical condition. An Energy Cepstral Coefficient (ECC) and filter envelope for features was proposed, and an Artificial Neural Network model and HMM was used for classification. For a Signal to Noise Ratio (SNR) of 15, the cough detection rate is about 91%. S. Matos [14] proposed an HMM-based classification method using MFCC to classify cough sounds with an accuracy of about 82%.

The remainder of this paper is structured as follows. Section 2 describes the segmentation method and the features of the snoring and coughing detection mechanism. Background noise is used to calculate a threshold to segment independent events, the gradient of the banded spectral magnitude sum is used to identify snore features, and MFCC [15] is used to identify cough features. In addition, Section 2 describes the classification method, which uses a Support Vector Machine (SVM) for snore detection and HMM for cough detection. The number and frequency of coughs and snores is important for diagnosis and treatment of upper respiratory tract diseases. The experimental results and analysis are given in Section 3. Section 4 details conclusions and future proposals.

2 Detection Method of Snoring and Coughing

The section describes the audio recording instrument, sleep environment, data collection, and detection method for snoring and coughing.

2.1 Audio Recording Instrument

This study uses a recording pen with a highly sensitive microphone. The recording pen is a Sony ICD-UX513F, and the highly sensitive microphone is an Audio-Technica AT9942, as shown in Figure 3. The recording pen was less effective than expected, so a highly sensitive and directional microphone was used to record sounds. The sample rate is 44100, and mono recordings are used.

Figure 3. Recording instrument

(a) Sony ICD-UX513F recording pen; (b) Audio-Technica AT9942 highly sensitive and directional microphone

2.2 Sleep Environment and Data Collection

Taichung Veterans General Hospital was the location of this study. Subjects presented with OSA and coughing. The audio signals from the subjects were collected at night in Taichung Veterans General Hospital Sleep Center. The sleep environment is shown in Figure 4. Fifteen subjects participated in this study. All subjects signed a consent form. The age, sex, Body Mass Index (BMI), and Apnea-Hypopnea Index (AHI) details for the subjects are shown in Table 1.

Figure 4. Sleep environment

Table 1. Subject information

	Age	BMI	AHI
Men (9)	54.4±12.7	29.3±4.5	44.8±21.8
Women (6)	44.6±9.3	26.9±3.7	14.9±10
All (15)	50.3±12.1	28.4±4.2	32.9±23.2

2.3 Algorithm

Figure 5 shows the flowchart for the proposed algorithm. At the beginning of the process, the sound data from subjects was segmented into independent events, features of snoring and coughing were recorded individually, and then snoring and coughing were detected. Independent events were then classified as snores, coughs, and other noises. A decision layer was used for a positive snore and cough detection result.

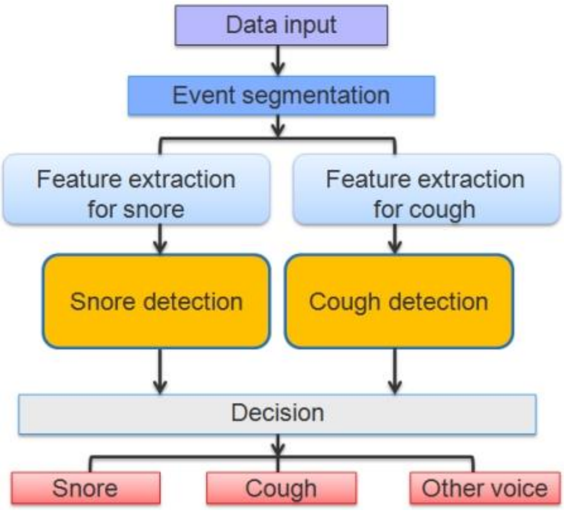


Figure 5. Flowchart

2.3.1 Event Segmentation

The Sleep Center is quiet, so background noise is consistent, as shown in Figure 6.

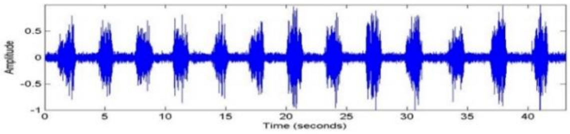


Figure 6. Audio recording data for a patient

Energy is conventionally used to determine the boundaries for sound activity. Energy is defined as:

$$X_a = \sum_{n=(a-1)\cdot r+1}^{r\cdot a} S^2(n), a = 1,2,\ldots,m, \quad \textbf{(1)}$$

where S is amplitude, r is the frame size, and a is the frame index for all audio signals. Energy is calculated to increase the difference in amplitude and for smoothing.

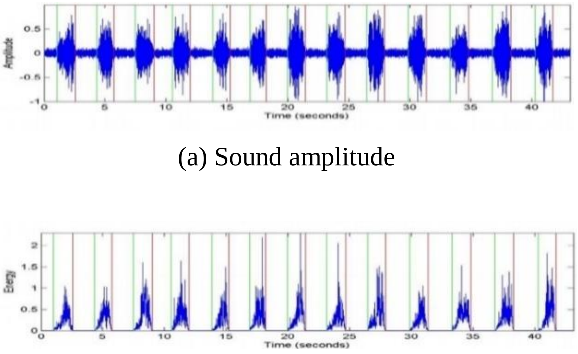
Background noise was consistent in the sleep environment, and the audio signal begins with background noise. Initially, the threshold for audio signal energy computation was defined as:

$$T_s = \mu + \alpha \cdot \sigma$$

$$\mu = \frac{1}{h}\sum_{k=1}^{h} X_k$$

$$\sigma = \sqrt{\frac{1}{h}\sum_{k=1}^{h}(X_k - \mu)^2}, \quad \textbf{(2)}$$

where μ is the average, σ is the standard deviation, and h is the frame number for the background noise. The control chart [16] indicates that the signal value is greater than a value which is the average plus the triple standard deviation, so α is set as 3 in the proposed method. An event occurs when the energy is greater than Ts. The segmentation results are shown in Figure 7, wherein the green lines represent the beginning of an event, and red lines mark the end of an event.



(a) Sound amplitude

(b) Sound energy

Figure 7. Segmentation results

2.3.2 Feature Extraction for Snores

Figure 8 shows the flowchart for feature extraction from snores. First, each independent event in the time domain is transformed into the frequency domain signal using a differential and Fourier transform. Next, the banded spectral magnitude sum gradient is calculated and normalization is performed. The dimension is then reduced by principal component analysis. Finally, the feature vector is output.

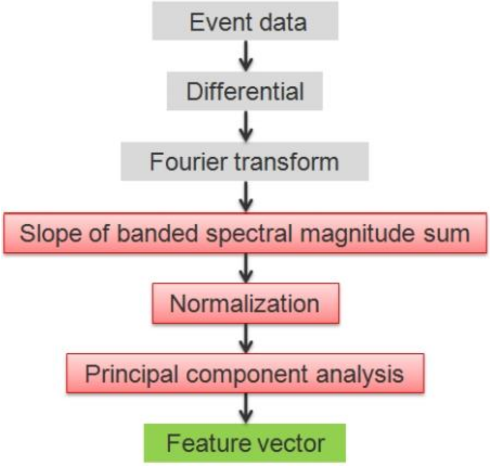

Event data
Differential
Fourier transform
Slope of banded spectral magnitude sum
Normalization
Principal component analysis
Feature vector

Figure 8. The flowchart for feature extraction from snores

2.3.2.1 Differential and Fourier Transform

It is difficult to extract features from audio signals in the time domain. The audio signal is transformed from the time domain into the frequency domain. However, background noise can affect the spectrum so a differential [15] is used to reduce the effect as follows:

$$\hat{S}(n) = S(n) - a \cdot S(n-1), \qquad \textbf{(3)}$$

where a is between 0.9 and 1.

The audio signal is then transformed from the time domain into the frequency domain using a Fourier transform, which is defined as:

$$F(k) = \sum_{n=0}^{N-1} \hat{S}(n) e^{\frac{-2jk\pi n}{N}}, k = 0,1,\dots,N-1, \textbf{(4)}$$

where N is the number of signal in a time window. Figure 9 shows the spectrum for an audio signal subject to a Fourier transform. The x-axis is the frequency and the y-axis is the magnitude for each frequency.

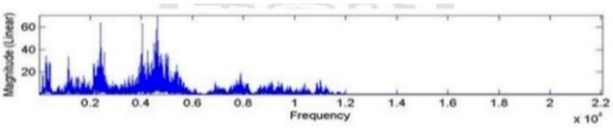
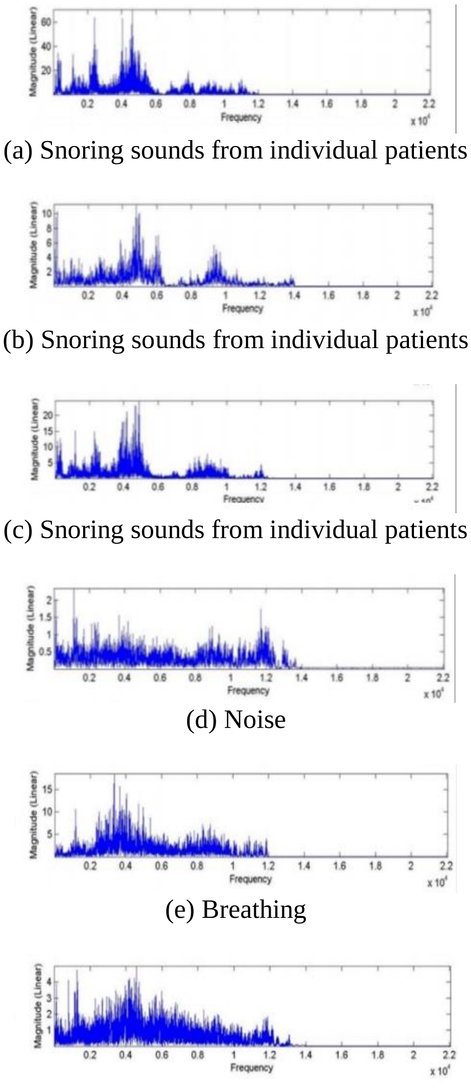


Figure 9. The spectrum for an audio signal

2.3.2.2 Gradient of the Banded Spectral Magnitude Sum

The spectrum determines the consistency of the snoring sound. The different sound spectrums from the database are shown in Figure 10. Snoring sounds from different patients are dissimilar, but there are similarities between some spectral bands. Some spectral bands are high and some are low. The banded spectral magnitude sum gradient shows the consistency of snoring sounds and the difference in non-snoring sounds.

The snoring sound spectrum is used to determine the difference between the peaks of each band. If the band size is too small, the computational complexity is high. Sound consistency and difference are not shown. The band size from smaller peaks is used, 100Hz, as shown in Figure 11. The band size is 100Hz.



(a) Snoring sounds from individual patients

(b) Snoring sounds from individual patients

(c) Snoring sounds from individual patients

(d) Noise

(e) Breathing

(f) Knocks

Figure 10. Spectra

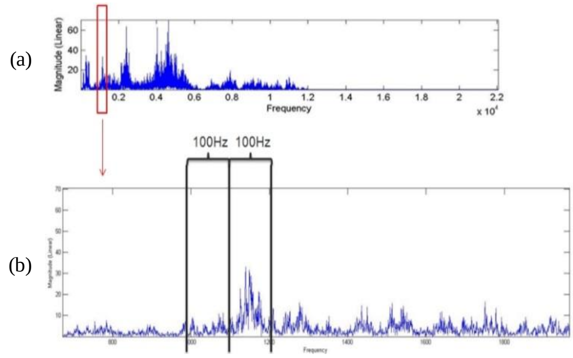


Figure 11. (a) Sound spectrum (b) The smaller peak diagram

The magnitude of the bands is then summed individually as:

$$BS(i) = \sum_{f=U\cdot(i-1)}^{U\cdot i} |S(f)|, i = 1,2,\dots,L$$

$$L = \left[\frac{22050}{U}\right], \qquad \textbf{(5)}$$

where *S*(*f*) is the magnitude of frequency *f* of the spectrum, *L* is the number of bands, and *U* is the band size. The gradient of each band is calculated as:

$$D(i) = \frac{\sum_{\tau=-M}^{M} BS(i+\tau)\cdot\tau}{\sum_{\tau=-M}^{M}\tau^2}, \quad \textbf{(6)}$$

where *M* will affect the accuracy of detection result and the detail discussion is discussed in Section 3. If *D* is positive, and then *BS* increases.

2.3.2.3 Normalization

Snoring sounds are consistent, as shown in Figure 10, but the volume of snoring sounds is not constant so the magnitude of the snoring spectrum changes. The results are normalized to identify snoring characteristics. A *Z*-score is used to normalize the gradient of the banded spectral magnitude sum for each snore. The *Z*-score is defined as:

$$Z_i = \frac{D(i)-\mu}{\sigma}, \quad \textbf{(7)}$$

where μ is the average of *D* and σ is the standard deviation of *D*, which is defined as:

$$\mu = \frac{1}{L}\sum_{i=1}^{L} D(i)$$

$$\sigma = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(D(i)-\mu)^2}. \quad \textbf{(8)}$$

2.3.2.4 Principal Component Analysis

Principal Component Analysis (PCA) [17-20] is used to reduce the dimensions of the feature vector, as shown in Figure 12. PCA uses linear projection for transformation, uses a few dimensions to represent all dimensions and then keeps the original characteristics of variation.

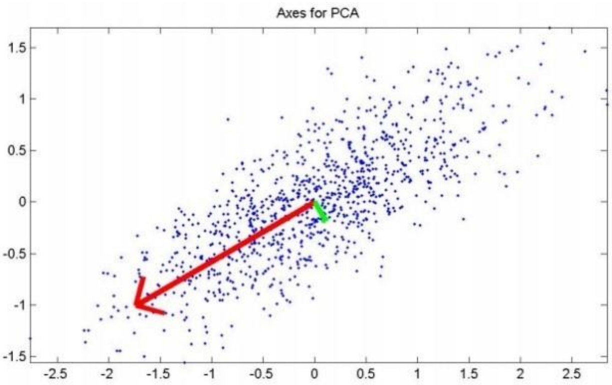


Figure 12. Diagram of principal component analysis

2.3.3 Feature Extraction for Coughs

Features are extracted for coughs using the Mel-Frequency Cepstral Coefficient (MFCC) [15], which accounts for the sensitivity of the human ear to different frequencies, so this parameter is used for speech recognition and speaker recognition [21]. The procedure for the MFCC is shown in Figure 13.

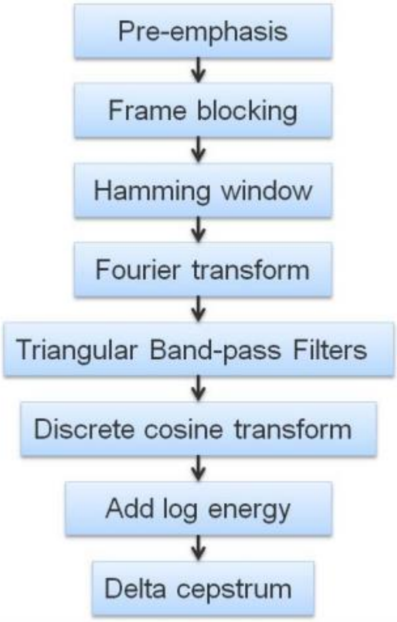


Figure 13. The procedure for the MFCC

The first step for the MFCC is pre-emphasis, which is defined as:

$$\hat{S}(n) = S(n) - a\cdot S(n-1), \quad \textbf{(9)}$$

where *a* has a value between 0.9 and 1. There is always noise in the air so there is more low-frequency than high-frequency energy. The accuracy of recognition is unaffected by this condition. Pre-emphasis reduces low-frequency energy relative to high-frequency energy. High- frequency formants are highlighted [22].

The next step is frame blocking. The shortest cough sounds in the sound database are between 0.25 and 0.3 seconds, and the frame number is at least 10. Thus, the frame size is about 25 ms. The sample rate is 44100 for this database, so the frame size is 1024 sample points, as shown in Figure 14.

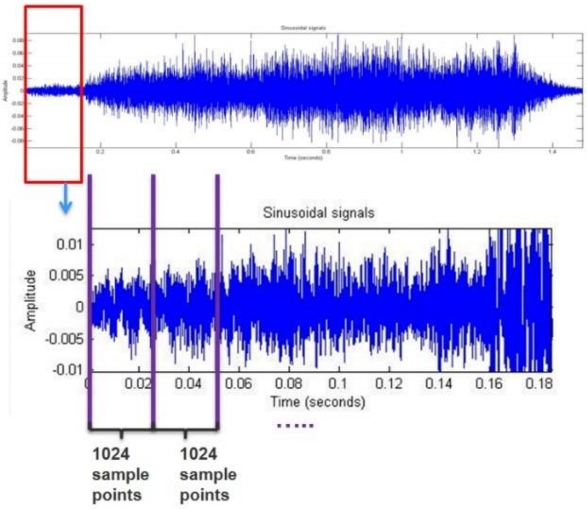


Figure 14. Diagram for frame blocking

Each frame is then multiplied by the hamming window, as follows:

$$S'(n) = \hat{S}(n)\cdot W(n), \quad \textbf{(10)}$$

where $W(n)$ is hamming window, which is defined as:

$$W(n,\rho) = (1-\rho) - \rho \cdot \cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \le n \le N-1, \tag{11}$$

where ρ is a value that represents different hamming windows, as shown in Figure 15. The value of ρ is 0.46.

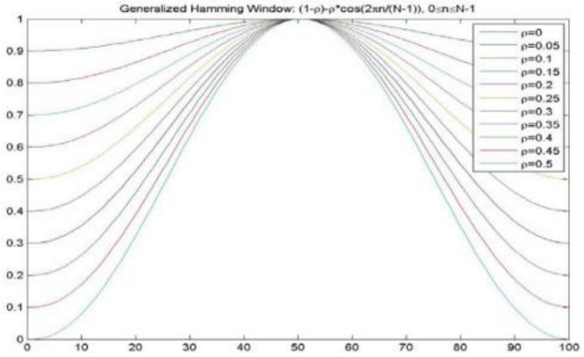


Figure 15. Diagram of the hamming window

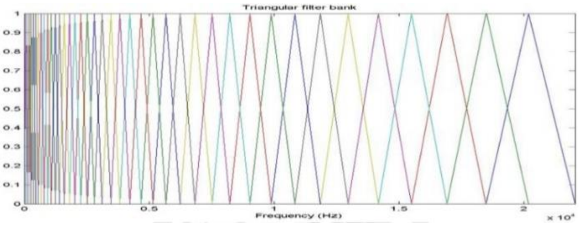


Figure 16. Diagram of a triangular filter

The next step is the Fourier transform, defined in Equation (4). It is difficult to extract features from the audio signal in the time domain so it is transformed to the frequency domain using a Fourier transform for signal processing. The same sounds are consistent in the spectrum. The result is multiplied by the hamming window to prevent an erroneous spectral magnitude, which leads to a calculation error due to discontinuity between frames.

The spectral magnitude through a 40 triangular band-pass filter [23] is shown in Figure 16. The logarithm of each triangular band-pass filter output is calculated. The 40 triangular band-pass filters are evenly distributed in the Mel-frequency. The relationship between the Mel-frequency and the normal frequency is defined as:

$$mel(f) = 2595 \cdot \log_{10}(1 + \frac{f}{100}), \tag{12}$$

where f is the normal frequency. The relationship between the Mel-frequency and the normal frequency diagram is shown in Figure 17. The human ear is more sensitive to low frequency energy than high frequency energy. The increase in sensitivity is logarithmic.

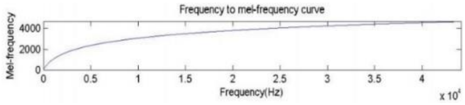


Figure 17. Diagram of the mel-frequency

A discrete cosine transform, which is a powerful transform to extract proper features, is then used [24]. The Mel-frequency cepstral parameter is obtained using the 40 logarithms through a discrete cosine transform. The discrete cosine transform is defined as:

$$C_m = \sum_{k=1}^{N} \cos\left(\frac{m \cdot (k-0.5) \cdot \pi}{N}\right) \cdot E_k \quad m = 1,2,\ldots,L, \tag{13}$$

where N is the number of triangular band-pass filters, E_k is the logarithm of the triangular band-pass filter, m is the number of mel-scale cepstral coefficients, and L is 12. The discrete cosine transform transforms the frequency domain into time-domain signal. This is the cepstrum calculation.

This parameter is called the Mel-frequency cepstral coefficient. A frame logarithmic energy is added as:

$$LogEnergy = 10 \cdot \log_{10} FrameEnergy. \tag{14}$$

There are 13-dimensional parameter vectors but in practice, the delta cepstrum is added to show the Mel-frequency's cepstrum dynamic variation in time. This is defined as:

$$\Delta C_m = \frac{\sum_{\tau=-M}^{M} C_m(t+\tau) \cdot \tau}{\sum_{\tau=-M}^{M} \tau^2}, \tag{15}$$

where M is 2. The delta cepstrum is added, so the Mel-frequency cepstral coefficient is a 39-dimensional parameter vector.

2.4 Sound Detection

This study uses two methods for classification: a Support Vector Machine (SVM) [25-26] and a HMM.

2.4.1 Support Vector Machine

A SVM separates the worst case for two groups and determines the optimal separating hyperplane. Two groups can be separated once the worst case is separated, as shown in Figure 18.

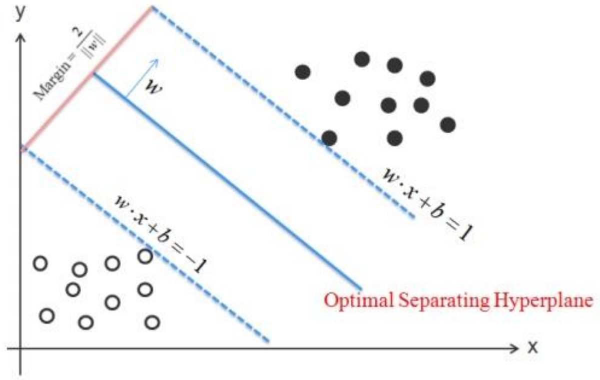


Figure 18. Diagram of a SVM

2.4.2 Hidden Markov Model

A HMM is commonly used in speech recognition. A HMM is different from other classifiers because it cannot be applied to a time-varying signal, but a HMM can be applied.

Coughing is a time-varying signal so a HMM is used as a classifier.

A diagram of a HMM is shown in Figure 19. There are three parameters. The first parameter is the initial probability, which is defined as:

$$\pi_i = P(q_0 = S_i), \tag{16}$$

where π_i is the initial probability in state i. The second parameter is the transition probability, which is defined as:

$$a_{ij} = P(q_{k+1} = S_j | q_k = S_i), \tag{17}$$

where a_{ij} is the probability of state i transferring to state j. The last parameter is the observation probability, which is defined as:

$$b_j(O_t) = P(O_t | q_k = S_i), \tag{18}$$

where $b_j(O_t)$ is the probability of observation in state j.

A HMM only gives the observation sequence, not the state sequence. Therefore, the transition and observation probability is calculated using the Baum-Welch algorithm [27] and the state sequence is calculated using the Viterbi algorithm [28].

The Baum-Welch algorithm estimates the initial probability of the HMM. By using the observation sequence, the transition probability and observation probability are constantly updated until convergence, as follows:

$$\bar{\pi}_i = \gamma_1(i)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T} \gamma_t(i)}$$

$$\bar{b}_j(v_k) = \frac{\sum_{t=1\, s.t.\, O_t = v_k}^{T-1} \gamma_t(i,j)}{\sum_{t=1}^{T} \gamma_t(i)}, \tag{19}$$

where γ_t (i) is the all path probability sum through state i at time t, as shown in Figure 20. ξ_t (i, j) is the all path probability sum through state i to state j, as shown in Figure 21.

The Viterbi algorithm is shown in Figure 22. It determines the most excellent probability path if there is an observation sequence but no state sequence. The initial probability is defined as:

$$\delta_1(j) = \pi_j \cdot b_j(O_1). \tag{20}$$

The greatest probability path at time t is then determined and the greatest path at time t+1 is calculated. The formula is:

$$\delta_{t+1}(j) = \left[\max_i\left(\delta_t(i) \cdot a_{ij}\right)\right] \cdot b_j(O_{t+1}). \tag{21}$$

The excellent path for this HMM is then calculated. This study uses a HMM to train and test coughs. The following describes the training and testing process.

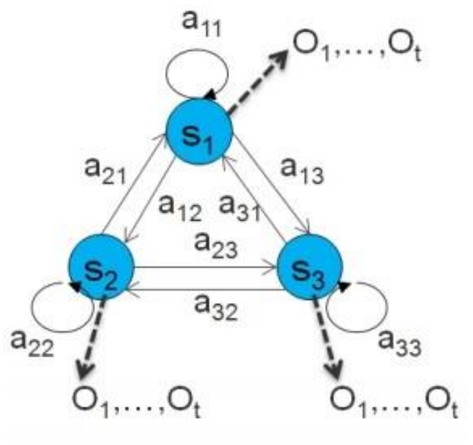


Figure 19. Diagram of a HMM

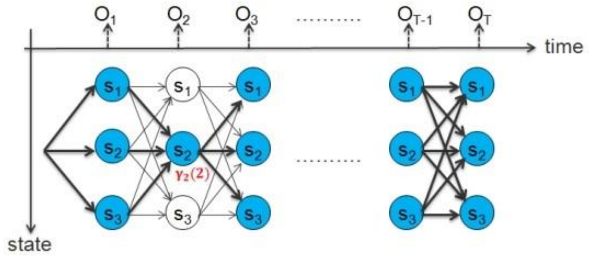


Figure 20. Diagram for γ_t (i)

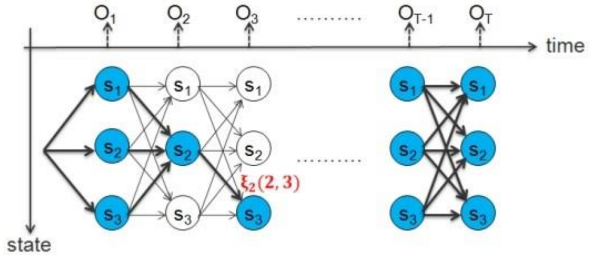


Figure 21. Diagram for ξ_t (i, j)

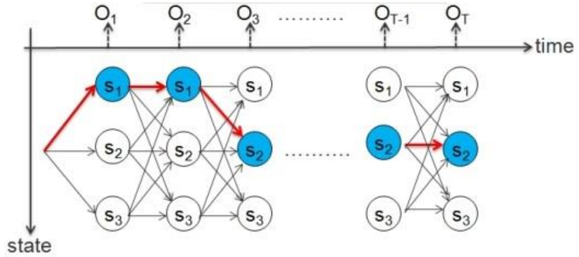


Figure 22. Diagram of the Viterbi algorithm

2.4.2.1 Hidden Markov Model – Training Phase

When the sound feature vectors are identified, observations are sorted using vector quantization to allow similar feature vectors to converge in the same group, which reduces the amount of computation. A k-means clustering algorithm [29-30] is used to quantify vectors as follows:

$$\underset{s}{\operatorname{argmin}} \sum_{i=1}^{M} \sum_{x_j \in S_i} \left\| x_j - \mu_i \right\|^2, \tag{22}$$

where $S = \{S_1, S_2, \ldots, S_M\}$, μ_i is the average of S_i, x_j is the feature vector, and M is the number of clusters. The k-means clustering algorithm is used to produce a codebook, which is $\{v_k|k = 1, 2, \ldots, M\}$. All feature vectors are then quantified using the codebook, as:

$$O_t = k \text{, if } d(x_t, v_k) < d(x_t, v_m), \\ \forall m \neq k \text{, then } x_t \in S_k, \quad \textbf{(23)}$$

where k is the index for the codebook and $d(x_t, v_k)$ is the distance between x_t and v_k.

When all observations are considered, the transition and observation probability is calculated using the Baum-Welch algorithm [27]. The state sequence is calculated using the Viterbi algorithm [28]. The HMM is then defined using an iterative process.

2.4.2.2 Hidden Markov Model – Testing Phase

To classify N types of sounds, N HMMs are created and the sound feature vectors are put into these N HMMs. The highest probability from these N HMMs is then determined. A cough HMM can be established in terms of cough detection, but a non-cough HMM cannot. A threshold is defined to differentiate between cough and non-cough events. A scoring mechanism is defined as:

$$score = \frac{b_{w1}(O_{R1}) + b_{w2}(O_{R2}) + \cdots + b_{wk}(O_{Rk})}{K}, \quad \textbf{(24)}$$

where K is the length of sound, O_R is the observation sequence of the sound, and wk is the best state sequence that is calculated using the Viterbi algorithm. The threshold is defined as:

$$threshold = \mu - \alpha \cdot \sigma, \quad \textbf{(25)}$$

where μ is the average of the cough training set score and σ is the standard deviation of the cough training set score. If the score is greater than the threshold, a cough is identified.

2.5 Decision

The snore and cough detection process requires a decision mechanism to determine all possible detection conditions, as shown in Figure 23. If snore detection results are positive but cough detection is negative, a snore is identified. If the snore detection results are negative but cough detection is positive, a cough is identified. If the snore detection results are negative and cough detection is negative, another noise is identified. If the result for snore detection is positive and cough detection is positive, the ratio of snore and cough sounds for the subject during sleep is assessed to determine whether a snore or cough has occurred.

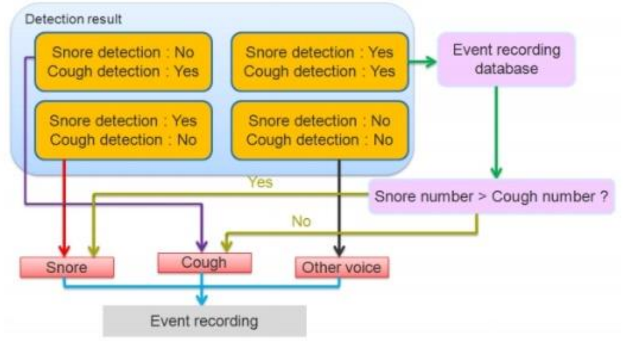


Figure 23. Decision mechanism

3 Experimental Results

3.1 Decision

The study uses Taichung Veterans General Hospital Department of Chest Medicine and Sleep Center. A database was collected for OSA and cough patients during a night's sleep. The total number of subjects is 15. The database is shown in Table 2.

Table 2. Sound database

Sounds	Number
Snore	22629
Cough	321
Breath	12976
Moan	532
Knock	375
Clear throat	111

3.2 Experimental Environment

This study uses Matlab and C++. The operating system is Microsoft Windows 7 (64-bit) and the development software is Matlab 2010a.

3.3 Results and Analysis

The experiment uses accuracy and sensitivity for detection validation. These are defined as:

$$Sensitivity(\%) = \frac{TP}{TP+FN} \cdot 100$$

$$Accuracy(\%) = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100, \quad \textbf{(26)}$$

where TP is a true positive, TN is a true negative, FP is a false positive, and FN is a false negative, as shown in Table 3 and Table 4.

Parameter validation uses k-fold cross-validation. One subject is omitted to demonstrate the detection accuracy for snores and coughs. For snore detection, the dimensions are reduced using PCA. The dimensional parameter validation is shown in Figure 24. If the dimension is 170, accuracy is constant, so 170 dimensions are used.

Table 3. Illustrations of TP, TN, FP and FN for snore detection

	Actual snore	Actual non-snore
Detected snore	TP	FP
Detected non-snore	FN	TN

Table 4. Illustration of TP, TN, FP and FN for cough detection

	Actual cough	Actual non-cough
Detected cough	TP	FP
Detected non-cough	FN	TN

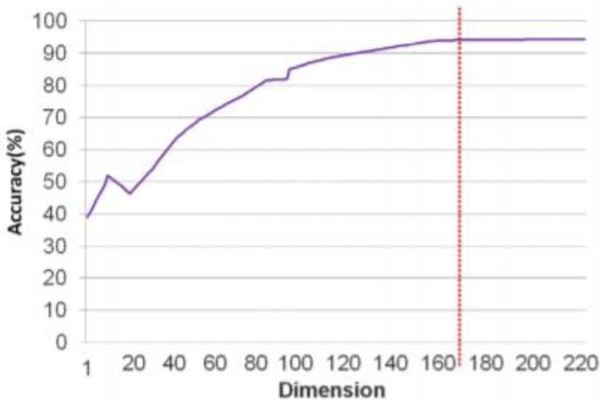


Figure 24. Dimensional parameter validation using PCA

In Equation (6), the *M* value affects the detection result, so the *M* value is validated, as shown in Figure 25. The accuracy is most outstanding if *M* is 2.

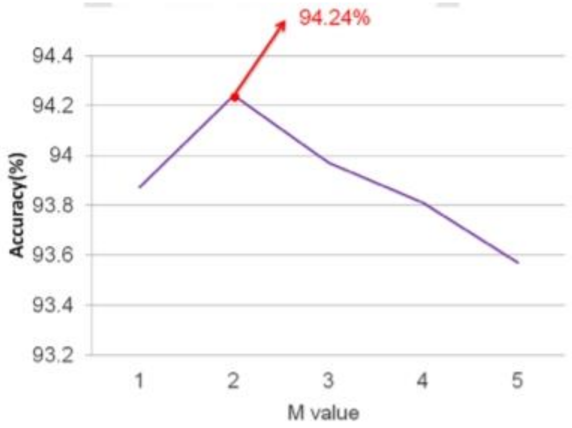


Figure 25. *M* value validation for Equation (6)

Table 5 shows the detection result for the proposed method. 2-fold, 3-fold, 5-fold, and 10-fold validation are used, and then the most excellent accuracy is 94.24%. One subject is omitted to determine the sensitivity of each patient using the proposed method, as shown in Table 6. The accuracy is 90% to 98%.

For cough detection, the clustering number must be validated in the codebook, and the state number of HMM and α value in Equation (25) must be computed. In the paper, 3-fold cross-validation is used, as shown in Figure 26, Figure 27, and Figure 28. The greatest accuracy is for 640 clusters, 5 states, and a value for *α* of 1.55.

One subject is omitted to determine the sensitivity of each patient using MFCC and HMM, as shown in Table 7. The accuracy is 88.9% to 94.4%.

Table 5. Snore detection accuracy using *k*-fold cross validation

	Accuracy (%)
2-fold	93.62%
3-fold	93.88%
5-fold	94.05%
10-fold	94.24%

Table 6. The sensitivity of snore detection

Patient No.	Snore number	Sensitivity (%)
#01	4483	95.95%
#03	2676	93.38%
#04	1223	90.59%
#05	2006	85.89%
#06	371	64.15%
#07	1991	75.34%
#08	4114	83.37%
#09	874	97.82%
#10	30	90.00%
#11	982	67.01%
#12	1512	89.15%
#14	1459	90.8%
#15	419	85.2%
#17	519	96.14%

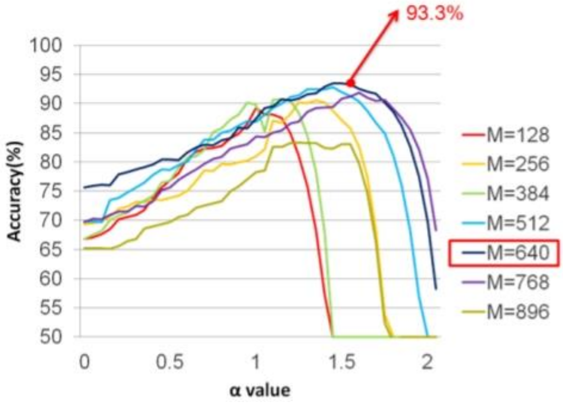


Figure 26. The accuracy of the number of clusters relative to α value for 3 states

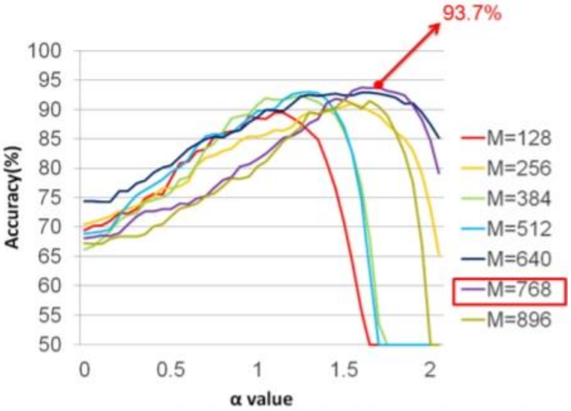


Figure 27. The accuracy of the number of clusters relative to *α* value for 4 states

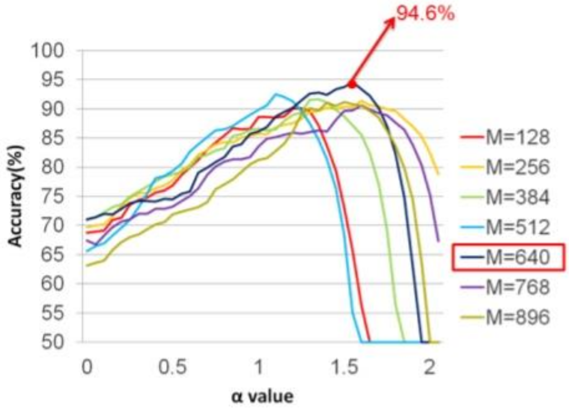


Figure 28. The accuracy of the number of clusters relative to the α value for 5 states

Table 7. The sensitivity of cough detection

Patient No.	Cough number	Sensitivity (%)
#01	19	89.4%
#03	4	75%
#06	12	83.3%
#07	7	57.1%
#08	2	100%
#10	90	94.4%
#11	1	100%
#13	20	80%
#14	144	91.7%
#15	4	75%
#17	18	88.9%

3.4 Discussion

Snoring is detected with an accuracy of 93% using k-fold cross validation but the accuracy has a wide range if one subject is omitted. The greatest value is 97.82% and the lowest is 64.15%. First possible reason is that the groups of subjects are different, so snoring sounds are also different. Second one is that the sound database is small.

Coughing is detected with an accuracy of 94.6% using *k*-fold cross-validation, but is not as accurate if one subject is omitted. The database is also small.

4 Conclusions and Future Work

4.1 Conclusions

Snores are detected using the gradient of the banded spectral magnitude sum for feature vectors, which shows the consistency of each sound spectrum. The principal component analysis was used to reduce the dimension and a SVM was used to classify snores and non-snores.

In terms of cough detection, coughs are time-variant so resemble speech. A MFCC was used to extract features and a HMM was used as a classifier, similar to speech recognition. The method produces good results.

4.2 Future Work

Each person’s snore is different so to increase accuracy, types of snoring sounds may be differentiated and the method and features improved. The sound database is also small in this study so the detection accuracy is not optimal. A more extensive sound database may increase accuracy.

References

[1] National Institutes of Health, Brain Basics: Understanding Sleep, NIH Publication, https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Understanding-Sleep

[2] S. Brand, R. Kirov, Sleep and its importance in adolescence and in common adolescent somatic and psychiatric conditions, *International Journal of General Medicine*, Vol. 4, pp. 425-442, June, 2011.

[3] G. J. Landry, T. Liu-Ambrose, Buying time: A rationale for examining the use of circadian rhythm and sleep interventions to delay progression of Mild Cognitive Impairment to Alzheimer’s disease, *Frontiers Aging Neuroscience*, Vol. 6, Article No. 325, December, 2014.

[4] J.-Z. Yan, B. Hu, H. Peng, H.-Y. Ma, W. Zhao, An Ubiquitous Sleep Quality Monitoring and Evaluation, *Journal of Internet Technology*, Vol. 12, No. 3, pp. 375-381, May, 2011.

[5] R. Heinzer, S. Vat, P. Marques-Vidal, H. Marti-Soler, D. Andries, N. Tobback, V. Mooser, M. Preisig, A. Malhotra, G. Waeber, P. Vollenweider, M. Tafti, J. Haba-Rubio, Prevalence of sleep-disordered breathing in the general population: the HypnoLaus study, *Lancet Respiratory Medicine*, Vol. 3, No. 4, pp. 310-318, April, 2015.

[6] P. Lévy, M. Kohler, W. T. McNicholas, F. Barbé, R. D. McEvoy, V. K. Somers, L. Lavie, J.-L. Pépin, Obstructive sleep apnoea syndrome, *Nature Reviews Disease Primers*, Vol. 1, No. 1, Article No. 15015, December, 2015.

[7] O. M. Bubu, A. G. Andrade, O. Q. Umasabor-Bubu, M. M. Hogan, A. D. Turner, M. J. de Leon, G. Ogedegbe, I. Ayappa, Jean-Louis G. Girardin, M. L. Jackson, A. W. Varga, R. S. Osorio, Obstructive sleep apnea, cognition and Alzheimer’s disease: a systematic review integrating three decades of multidisciplinary research, *Sleep Medicine Reviews*, Vol. 50, Article No. 101250, April, 2020.

[8] A. Romero-Corral, S. M. Caples, F. Lopez-Jimenez, V. K. Somers, Interactions between obesity and obstructive sleep apnea: implications for treatment, *Chest*, Vol. 137, No. 3, pp. 711-719, March, 2010.

[9] M. Cavusoglu, M. Kamasak, O. Eroğul, T. Çiloglu, Y. Serinagaoglu Dogrusoz, T. Akcam, An efficient method for snore/nonsnore classification of sleep sounds, *Physiological Measurement*, Vol. 28, No. 8, pp. 841-853, August, 2007.

[10] A. Azarbarzin, Z. M. K. Moussavi, Automatic and Unsupervised Snore Sound Extraction From Respiratory Sound Signals, *IEEE Transactions on Biomedical Engineering*, Vol. 58, No. 5, pp. 1156-1162, May, 2011.

[11] W. Duckitt, S. Tuomi, T. Niesler, Automatic detection, segmentation and assessment of snoring from ambient acoustic data, *Physiological Measurement*, Vol. 27, No. 10, pp. 1047-1056, October, 2006.

[12] A. S. Karunajeewa, U. R. Abeyratne, C. Hukins, Silence-breathing-snore classification from snore-related sounds, *Physiological Measurement*, Vol. 29, No. 2, pp. 227-243, February, 2008.

[13] S. Shin, T. Hashimoto, S. Hatano, Automatic Detection System for Cough Sounds as a Symptom of Abnormal Health Condition, *IEEE Transactions on Information Technology in Biomedicine*, Vol. 13, No. 4, pp. 486-493, July, 2009.

[14] S. Matos, S. S. Birring, I. D. Pavord, H. Evans, Detection of cough signals in continuous audio recordings using hidden Markov models, *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 6, pp. 1078-1083, June, 2006.

[15] Q. Mei, M. Gül, M. Boay, Indirect health monitoring of bridges using Mel-frequency cepstral coefficients and principal component analysis, *Mechanical Systems and Signal Processing*, Vol. 119, pp. 523-546, March, 2019.

[16] G. Suman, D. R. Prajapati, Control chart applications in healthcare: A literature review, *International Journal of Metrology and Quality Engineering*, Vol. 9, Article No. 5, May, 2018.

[17] Z.-G. Chen, H.-S. Kang, S.-R. Kim, Design of a New Efficient Hybrid System for Intrusion Detection Based on HSM Fuzzy Decision Tree, *Journal of Internet Technology*, Vol. 16, No. 5, pp. 885-891, September, 2015.

[18] S.-S. Weng, K.-Y. Chen, C.-Y. Li, A Geometric Mean-based DEMATEL Model for Evaluating the Critical Challenges of Spare Parts Planning, *Journal of Internet Technology*, Vol. 21, No. 1, pp. 121-133, January, 2020.

[19] J. S.-W. Wan, S.-D. Wang, Concept Drift Detection Based on Pre-Clustering and Statistical Testing, *Journal of Internet Technology*, Vol. 22, No. 2, pp. 465-472, March, 2021.

[20] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, T. Baker, Analysis of Dimensionality Reduction Techniques on Big Data, *IEEE Access*, Vol. 8, pp. 54776-54788, March, 2020.

[21] C. Ittichaichareon, S. Suksri, T. Yingthawornsuk, Speech recognition using MFCC, *Proceeding International Conference on Computer Graphics, Simulation and Modeling*, Pattaya, Thailand, 2012, pp. 135-138.

[22] L. E. Baum, T. Petrie, G. Soules, N. Weiss, A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, *The Annals of Mathematical Statistics*, Vol. 41, No. 1, pp. 164-171, February, 1970.

[23] S. C. Joshi, A. N. Cheeran, MATLAB Based Feature Extraction Using Mel Frequency Cepstrum Coefficients for Automatic Speech Recognition, *International Journal of Science Engineering and Technology Research* (*IJSETR*), Vol. 3, No. 6, pp. 1820-1823, June, 2014.

[24] S. Gupta, N. Dhanda, Audio Steganography Using Discrete Wavelet Transformation (DWT) & Discrete Cosine Transformation (DCT), *IOSR Journal of Computer Engineering*, Vol. 17, No. 2, pp. 32-44, March-April, 2015.

[25] S. Wang, Z. Tang, S. Li, Design and Implementation of an Audio Classification System Based on SVM, *Procedia Engineering*, Vol. 15, pp. 4031-4035, 2011.

[26] F. Rong, Audio Classification Method Based on Machine Learning, *International Conference on Intelligent Transportation, Big Data & Smart City* (*ICITBS*), Changsha, China, 2016, pp. 81-84.

[27] P. M. Baggenstoss, A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces, *IEEE Transactions on Speech and Audio Processing*, Vol. 9, No. 4, pp. 411-416, May, 2001.

[28] Q. Wang, L. Wei, R. A. Kennedy, Iterative Viterbi decoding, trellis shaping, and multilevel structure for high-rate parity-concatenated TCM, *IEEE Transactions on Communications*, Vol. 50, No. 1, pp. 48-55, January, 2002.

[29] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, A. Y. Wu, An efficient k-means clustering algorithm: analysis and implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 881-892, July, 2002.

[30] P. He, S. Ma, W. Li, Efficient Barrage Video Recommendation Algorithm Based on Convolutional and Recursive Neural Network, *Journal of Internet Technology*, Vol. 22, No. 6, pp. 1241-1251, November, 2021.

Biographies

Chun-Hung Yang is an Assistant Professor in the Department of Electronic Engineering, Southern Taiwan University of Science and Technology (STUST). His main research interests include artificial Intelligence (AI) in medicine and model-based design (MBD) for digitally-controlled power converters and motor drivers.

Yung-Ming Kuo received the Ph.D. degree in electrical engineering from National Cheng Kung University, Taiwan, in 2010. He is now an Assistant Professor in the Dept. of Electronic Engineering, National Formosa University, Taiwan. His research interests include deep learning, medical image analysis, computer vision, pattern recognition and video-based behavior analysis.

I-Chun Chen obtained her doctoral degree in literature at Kanazawa University in Japan in 2015. She is an Assistant Professor at National Formosa University, Yunlin, Taiwan. Her area and interest in research are in Chinese dialects and grammar, and Linguistics.

Fan-Min Lin received the Master of Science from the Institute of Computer and Communication Engineering, National Cheng Kung University, Tainan, Taiwan, R.O.C., in 2013. His main research interests include artificial Intelligence (AI), machine learning, digital signal processing, and audio recognition.

Pau-Choo Chung received the Ph.D. degree in electrical engineering from Texas Tech University, USA, in 1991. She then joined the Department of Electrical Engineering, National Cheng Kung University, Taiwan, and has become a full professor in 1996. Her research interests include computational intelligence, image analysis, video analysis, and pattern recognition.