A Novel Snore Detection and Suppression Method for a Flexible Patch With MEMS Microphone and Accelerometer

Chunhua He, Jiewen Tan, Xuelei Jian, Guangxiong Zhong, Heng Wu, Lianglun Cheng, and Juze Lin

***Abstract*—Sleep apnea impacts more and more people all over the world, and obstructive sleep apnea of which is the most frequent. Hence, research on snoring detection and related suppression methods is extremely urgent. In this article, a novel low-cost flexible patch with MEMS microphone and accelerometer is developed to detect snore event and sleeping posture, and a small vibration motor embedded in the patch is designed to suppress snoring. Theoretical analyses of short-time energy, piecewise average filtering (PAF), and Mel-frequency cepstral coefficients (MFCCs) processing are described in detail, and the improved MFCCs are put forward and used as the input of the convolutional neural network (CNN). Furthermore, the snore recognition method based on the combination of similarity analysis and CNN analysis is presented, followed by the snoring suppression method. Experimental results demonstrate that the main features of the sound signals can be extracted effectively by PAF and MFCCs processing, and the data compression ratio is about 99.41%. Besides, the locations of the eigenvectors can be found accurately based on short-time energy analysis. The numbers of high similarity of snoring signals within 30 s are larger than 3, while those of non-snoring signals are often less than 3. If the preliminary screening with similarity analysis is passed, CNN analysis will be conducted to judge whether there are snoring events. The accuracy of snore recognition with CNN analysis is calculated to be as high as 99.25%. Finally, the average snoring time measured by the smart patch with snoring suppression is reduced to 15 from 135 min, which indicates that the proposed snore recognition and suppression methods are effective.**

***Index Terms*—Artificial intelligence, flexible patch, MEMS microphone and accelerometer, sleep apnea, snore detection and suppression.**

Manuscript received 13 April 2022; revised 30 July 2022; accepted 11 August 2022. Date of publication 16 August 2022; date of current version 7 December 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62104047, Grant 62173098, and Grant U2001201. *(Corresponding authors: Heng Wu; Juze Lin.)*

Chunhua He and Lianglun Cheng are with the School of Computer, Guangdong University of Technology, Guangzhou 510000, China (e-mail: hechunhua@pku.edu.cn; llcheng@gdut.edu.cn).

Jiewen Tan, Xuelei Jian, and Guangxiong Zhong are with the Research and Development Department, Guangzhou 37 Degree Smart Home Company Ltd., Guangzhou 510000, China (e-mail: tjw@37smarthome.com; jianxuelei@37smarthome.com; zhongguangxiong@37smarthome.com).

Heng Wu is with the School of Automation, Guangdong University of Technology, Guangzhou 510000, China (e-mail: heng.wu@foxmail.com).

Juze Lin is with the Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510000, China (e-mail: 631753613@qq.com).

Digital Object Identifier 10.1109/JIOT.2022.3199085

I. Introduction

SLEEP apnea impacts about one billion people all around the world, if untreated in time, the risks of stroke and cardiovascular mortality will be increased by more than twice [1], [2]. The symptoms of sleep apnea are snoring, fatigue, and so on, while the main types of sleep apnea are obstructive sleep apnea (OSA), central sleep apnea, and complex sleep apnea. Among these OSA is the most frequent and can be treated or suppressed by correct sleeping posture [3]. OSA is one of the most common sleep disorders and is characterized by the frequent partial or complete collapse of the upper airway during sleep. Identifying the severity of airway obstruction is of great importance for successful treatment. The apnea hypopnea index (AHI), defined as the total number of full and partial upper airway obstructions per hour, is widely used to diagnose and characterize the severity of OSA. Patients are categorized according to their AHI values as follows: normal (AHI $<$ 5); mild (5 $\leq$ AHI $<$ 15); moderate (15 $\leq$ AHI $<$ 30); and severe (AHI $\geq$ 30). Snoring may be episodic, positional, continuous, or alternating, whereas in OSA syndrome, the snoring events only occur between successive respiratory obstructive events [4]. In mild snoring, the episodic pattern is the most frequent, whereas in moderate and severe snoring, the continuous snoring pattern occurs in most cases. Therefore, the techniques for accurate snoring detection and relevant treatment are very urgent, and related research has been paid more and more attentions.

In terms of the snore detection system, a variety of monitoring devices have been proposed in recent years. The Gold Standard for sleep-disordered breathing (SDB) diagnostics is done by using polysomnography (PSG). However, it is uncomfortable to wear a serial of sampling electrodes. Besides, it is time consuming and very expensive [5], which makes it not suitable for OSA screening at home. The American Association of Sleep Medicine (AASM) recommends three methods to detect snoring events: 1) a piezoelectric sensor placed on the neck; 2) an acoustic sensor, such as microphone on the trachea, chest, or near the patient's bed; and 3) nasal prongs connected to the nasal pressure sensor [6]. The piezoelectric sensor or stretchable strain sensor placed on the neck is uncomfortable to use, and the movement or sweat may affect the detection accuracy [7], [8]. Similarly, the nasal pressure sensor is also uncomfortable to use, especially for sleeping detection. Pressure monitoring-based mattress

type sensors, such as the electromechanical film transducer, static-charge-sensitive bed, polyvinylidene fluoride film, and balancing tube air-mattress, can be used for SDB diagnosis [9], [10], [11], [12], whereas the accuracy is not high enough. The RR intervals of the electrocardiography (ECG) or photoplethysmography (PPG) signals of smart watches can be utilized to detect the snore event [13], [14], [15], but the detection accuracy may be affected by the cardiovascular disease.

Recently, noncontact OSA detection methods are prevalent. Ultra wideband (UWB) and radar wave techniques have been applied to noncontact OSA screening [16], [17]. Nevertheless, the hardware cost is high, and the recognition precision is apt to be affected by the surroundings. Microphone-based sound analysis is another noncontact OSA screening method. Microphone is widely utilized to monitor the snoring events [18], [19], [20], [21], since it is small and cheap to be integrated into smartphone, pillow, or medical device. According to the feedback of microphone's signal analysis, the height of a smart pillow is adjusted automatically or a vibration motor is turned on so that the sleeper is woken up [22], [23]. However, the control system of the smart pillow is very complex and high cost, in addition, it is uncomfortable and nonportable for some users. Microphone in a smartphone can be used to monitor snoring, breathing, and other sounds [24], [25], and the snore event can be recognized since the acoustic spectrum of snore is different from those of other sounds. However, smartphone cannot intervene in the snoring therapy. Hence, microphone sensor can be adopted, but the detection system should be improved since comfortability and feedback control are very important.

Except for the snore detection system, the recognition methods are also very important. Nonspeech sounds, such as coughs, snores, crackles, wheezes, and asthmas, are associated with different respiratory diseases [26], [27], [28]. For instance, snores are related to sleep disorders, and coughs may be related to Coronavirus Disease-2019 (Covid-19) [29], [30], [31]. Therefore, in order to distinguish the snores from other sounds, pattern recognition is essential. Diverse recognition methods have been presented these years, among them, machine learning methods are prevalent. Either for image processing [32], [33] or for sound processing [34], [35], there are mainly two steps for machine learning, namely, feature extraction and modeling. For feature extraction, many methods have been proposed these years, such as principal component analysis (PCA), zero crossing rate, energy, entropy of energy, spectral centroid, spectral spread, spectral entropy, spectral flux [35], [36], Mel-frequency cepstral coefficients (MFCCs) [37], [38], [39], empirical mode decomposition, discrete wavelet transform [40], de-noising autoencoder technique, Gamma-tone frequency cepstral coefficients, and improved MFCCs [41], [42]. Among them, MFCCs are proved to be the most effective and widely used as the eigenvector, and some improvements about MFCCs are gradually presented.

For modeling, a variety of methods have been presented these years, such as support vector machine [43], random forest [44], attention BiLSTM network [45], recurrent neural network (RNN) [46], ResNet, ResNet-18, Resnet-50, long short-term memory (LSTM), multibranch deep learning network, and ensemble deep learning network [47]. Besides, some combination models have been proposed as well, such as convolutional neural network (CNN) followed by RNN [48], and deep neural network (DNN) combined with LSTM [49]. All these AI-based methods mentioned above have achieved a detection accuracy of about 95%, however, the implementation of these algorithms is very complicated and high cost. These methods can be hardly realized in the local microprogrammed control unit (MCU), which restricts the real-time application of snore detection at home. Even if these algorithms can be implemented in the cloud servers, remote data transmission and processing are very time consuming, high cost, and unreliable. Therefore, edge computing is inevitable for Internet of Things (IoT) products. Although some lightweight convolutional networks have been proposed [50], [51], [52], the architectures are still complex. FPGA device can be used to accomplish the AI (such as CNN-LSTM) processor, but the cost and power consumption are very high, so it is not suitable for wearable devices [53]. On the other hand, except for machine learning, similarity analysis can be also applied to the sound analysis. It is simple to realize, but the recognition accuracy is not high [54], [55]. Thereby, considering the cost and IoT edge computing, the combination of machine learning and similarity analysis may be better.

In our previous work, a novel smart flexible sleep monitoring belt with MEMS triaxial accelerometer and pressure sensor has been developed to detect vital signs, snore events, and sleep stages [56]. Although the snore events can be recognized, they cannot be suppressed in time. Hence, aiming at the analysis above, this article will propose a novel low-cost flexible patch with MEMS microphone and accelerometer to detect the snore event and sleeping posture, followed by snoring suppression by a small vibration motor. Meanwhile, statistics analysis and MFCCs will be applied for feature extraction, besides, machine learning and similarity analysis will be combined together to reduce the system's complexity of edge computing.

II. System Design for Flexible Patch

A smart flexible patch with a test circuit is shown in Fig. 1. It is mainly composed of an ordinary flexible fabric patch and a core circuit. The core circuit is packaged by a plastic box embedded in the patch. The block diagram of the core circuit system is depicted in Fig. 2, which mainly includes three subsystems, namely, power supply subsystem, processor subsystem, and sensor and actuator subsystem.

1) *Power Supply Subsystem:* Polymer lithium battery, DC/DC converter, and low-dropout regulator (LDO) are utilized to generate the appropriate supply voltages to the other subsystems.
2) *Processor Subsystem:* STM32F767 ARM is chosen as the core microprogrammed control unit (MCU), which has enough resources to achieve the snore recognition and feedback control algorithms. The USB interface

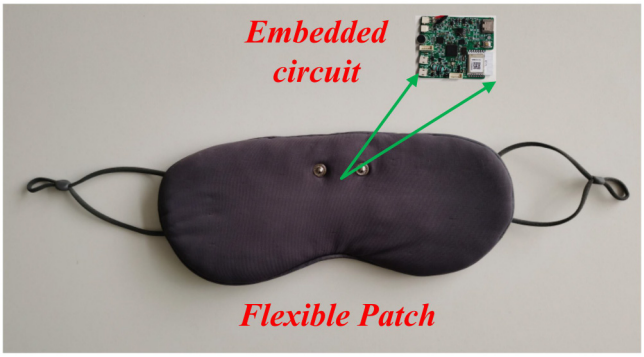


Fig. 1. Smart flexible patch with a test circuit.

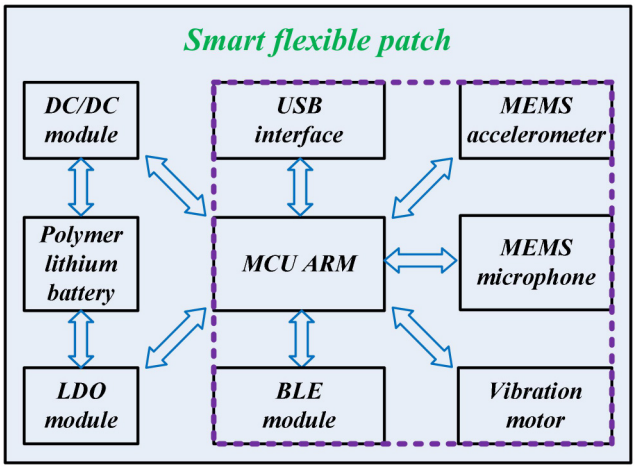


Fig. 2. Block diagram of the core circuit system for a smart flexible patch.

is utilized to connect with personal computer (PC), while the reserved Bluetooth (BLE) module is used to communicate with a smartphone for IoT application.

3) *Sensor and Actuator Subsystem:* The miniature and low-power MEMS microphone SPH0645LM4H is applied to detect the snore or ambient sound. Micropower three-axis MEMS accelerometer ADXL362 is adopted to acquire the movement and acceleration of the head. Thereby, the direction of the head and sleeping posture (such as supine or latericumbent) can be recognized and estimated. The small vibration motor is used to exert a little force to the head for suppressing snoring.

The flexible patch worn on the head is light and low cost. It can not only shut out the light but also detect the snore event and suppress snoring. Hence, based on this patch, a series of experiments can be conducted, as shown in Fig. 3, and the corresponding data are acquired for further processing and algorithms validation.

III. Snore Detection and Suppression Methods

Different sounds of varied scenarios can be acquired by the smart flexible patch, as shown in Fig. 4. It illustrates that there are great differences among the time-domain signals of snore, music, talk, and footsteps. As for snore, it has strong regularity. The duration of each snore sound is generally between 0.5 and 4 s, and the similarities between two adjacent snore signals are

Fig. 3. Flexible patch is worn on the head for sleep aid and suppressing snoring.

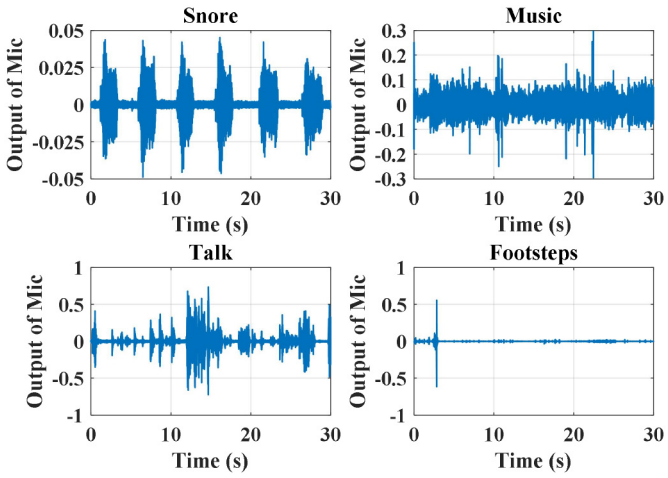


Fig. 4. Different sound signals acquired by the smart flexible patch.

relatively high. The characteristics of snore are so special that it ought to be possible to distinguish it from other sounds by further data processing and feature analysis. In this section, the snore recognition algorithm is put forward in detail, followed by a snoring suppression method.

A. Flowchart of Snore Recognition Algorithm

The flowchart of the snore recognition algorithm is depicted in Fig. 5, and the description is as follows.

1) *Data Acquisition and Preprocessing:* Before data processing for snore recognition, continuous data acquisition should be conducted first, and the sampling frequency is set to be 24 kHz. A large buffer in the MCU is created to save the collected sound data, and then online snore recognition analysis is performed every 30 s. In order to enhance the recognition accuracy, data preprocessing is essential, which includes pre-emphasis and framing.

2) *Time–Frequency Analysis and Processing:* First, calculate the short-time energy of each frame, and confirm the boundaries of every voice activity based on the

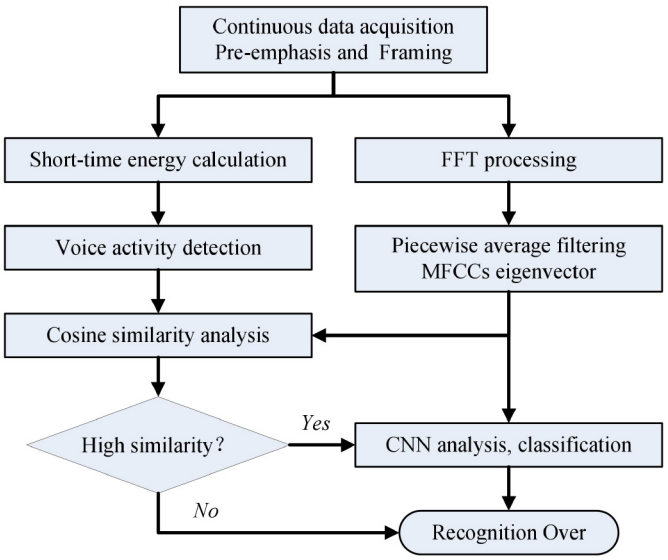


Fig. 5. Flowchart of the snore recognition algorithm.

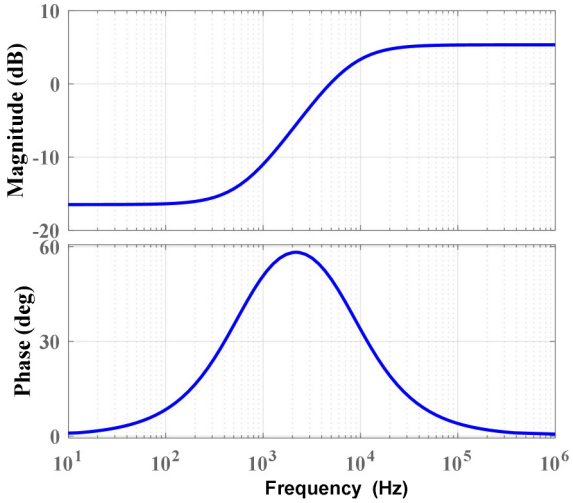


Fig. 6. Bode diagram of the pre-emphasis filter.

energy threshold. Then, the location of the feature vector of every sound envelope can be found. Meanwhile, fast Fourier transform (FFT) processing of each frame is done. Considering that the amount of data is large, dimension reduction processing of the FFT data is performed. Here, PAF and MFCCs processing are adopted. Hence, combining with the location of the feature vector found in the time-domain processing, the frequency-domain eigenvectors of PAF and MFCCs processing are obtained.

3) *Similarity and CNN Analyses:* Based on the multiple frequency-domain eigenvectors of PAF processing within 30 s, the similarities between two adjacent eigenvectors can be computed. If the similarities are high enough, then CNN analysis will be further conducted to judge whether there are snore events or not, and the frequency-domain eigenvectors of PAF and MFCCs processing are used as the input of the CNN; otherwise, it is judged to have no snore event within 30 s, and the snore recognition is over.

B. Data Acquisition and Preprocessing

1) Data Acquisition: The output of MEMS microphone is acquired by ARM MCU with the I^2S protocol. Considering that the frequency of human being's voice is often less than 12 kHz, thereby the sampling frequency is set to be 24 kHz, and the direct memory access (DMA) continuous sampling mode is adopted. The acquisition data within 30 s are stored in the flash, and a large buffer in the MCU is created to save the collected sound data. The principle of data storage is first-in–first-out (FIFO), which enables the sliding-window processing. Therefore, the data of the buffer is updated per about 0.042 ms by FIFO processing, and the data size of the buffer is always 720k. Based on these collected data, the snore recognition can be performed every 30 s or less. In this article, the processing duration is set to be 30 s to save the computing resource.

2) Data Preprocessing: In order to eliminate the effect caused by vocal cords and lips in the process of voice production, the high-frequency part of sound signal suppressed by the pronunciation system should be compensated. Thus, a first-order high-pass filter (HPF), i.e., pre-emphasis filter, is applied to enhance the high-frequency part and attenuate the low-frequency part at the same time. The time-domain formula of the HPF is as follows:

$$\begin{cases} y[i] = x[i], & i = 1 \\ y[i] = x[i] - k_p \times x[i-1], & 1 < i \end{cases} \tag{1}$$

where i stands for the ith moment, x is the input signal, and y is the output signal. k_p is the filter factor, determining the cut-off frequency. Accordingly, its transfer function in the discrete frequency domain (z domain) can be derived as

$$\frac{Y[z]}{X[z]} = \frac{z - k_p}{z}. \tag{2}$$

The equation of bilinear transformation between the discrete domain and the continuous domain is as follows:

$$z = \frac{2f_s + s}{2f_s - s} \tag{3}$$

where complex frequency s is equal to $j\omega$, and ω stands for the angular frequency. f_s represents the sampling rate. Thus, the transfer function in the continuous frequency domain (s domain) can be derived as

$$\frac{Y[s]}{X[s]} = \frac{2f_s(1 - k_p) + (1 + k_p)s}{2f_s + s}. \tag{4}$$

In this article, k_p is set to be 0.85, and f_s is set to be 24 kHz. Then, the bode diagram of the HPF can be depicted in Fig. 6. It demonstrates that the high-frequency part is amplified while the low-frequency part is suppressed, which matches the goal. Take the snoring signal as an example, the snoring signals with and without HPF filter processing are described in Fig. 7. It shows that the high-frequency part of the signal after filter is indeed magnified.

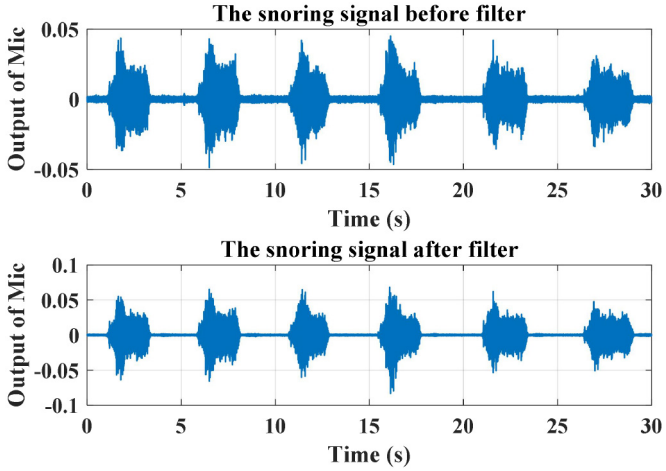


Fig. 7. Snoring signals before HPF filter and after HPF filter processing.

3) Framing Processing: After pre-emphasis processing, framing processing of the sound signal should be performed to reduce the computation complexity. Here, the size of the data per frame (N) is set to be 2048, whose corresponding time is 85.3 ms or so. Hence, there are about 352 frames within 30 s. After framing processing, time–frequency analysis and processing can be conducted.

C. Time–Frequency Analysis and Processing

1) Short-Time Energy Analysis: As for the signal of every frame, the total energy q can be yielded by calculating the sum of squares of the elements in the sound amplitude vector, as shown in

$$q = \sum_{i=1}^{N} y^2[i]. \tag{5}$$

Thus, there are 352 time-domain energies within 30 s, and they will form an energy vector ($\boldsymbol{Q}$). Each meaningful sound will last for a period of time, that is, the larger sound energy will appear in some consecutive frames. Therefore, in order to better distinguish a single sound signal, it is necessary to identify the positions of the energy envelope of a single sound signal. Here, a base line is set as the threshold for envelope positions recognition. The energy envelope of a single sound will cross the baseline twice, which yields the indexes of the starting and the ending frames of the sound envelope and confirms the starting and the ending positions of the single sound signal. Then, the duration (dT) of the sound envelope can be calculated according to the two indexes. In this article, the threshold is set dynamically according to the background noise, and the setting approach is as follows.

There are 352 energy values in the energy vector ($\boldsymbol{Q}$) within 30 s, if six values in succession are selected as a group, then 346 groups can be formed according to the sliding-window principle. Then, the standard deviation and average of each group, as well as their product, are calculated, as shown in (6). Next, the minimum is picked out from the 346 products and the corresponding group number is determined. Finally, the average energy of this group is multiplied by 4 and set as the threshold (THQ), as shown in (7), and then the base line is accordingly confirmed

$$\mathrm{PD}(i) = \mathrm{std}(Q(i{:}i+6)) \times \mathrm{mean}(Q(i{:}i+6)), 1 \le i \le 346 \tag{6}$$

$$\mathrm{THQ} = \min\ (\mathrm{PD}) \times 4 \tag{7}$$

where std(), mean(), and min() stand for the functions of standard deviation, average, and minimum. ***PD*** is the product vector. Given that the duration of snoring is generally larger than 0.5 s and less than 4 s, if dT is not within this range or the energy peak is tiny, it is judged that this sound envelope is invalid and further processing is unnecessary. Thereby, all the valid envelopes of the single sound signals are identified by this way, that is, the boundaries of every voice activity are clear.

There is often an energy peak located in the sound envelope, whose corresponding frame signal will possess the highest signal-to-noise ratio (SNR), and this frame signal can best represent the feature of this single sound signal. Therefore, based on the sound envelopes depicted above, the indexes corresponding to the energy peaks of all the valid sound envelopes are obtained, which are the locations of the feature vectors of the sound envelopes. These peak indexes make up a new vector (***IDP***), whose size is cnt.

2) Eigenvectors of PAF Processing: Followed by the time-domain processing, frequency-domain processing is also adopted to obtain the eigenvectors. Before FFT processing, the signal of each frame will be processed with a Hanning window function. Compared with the rectangular window function, the Hanning window can weaken the side-lobe size and spectrum leakage after FFT processing. Here, the formula of Hanning window function (W) is as follows:

$$W[i] = \left[1 - \cos\left(\frac{2\pi i}{L}\right)\right]/2, \quad 1 \le i \le L \tag{8}$$

where L is the length of the window. After processing with the Hanning window function, FFT spectrum analysis is conducted for the signal of each frame. The formula of the FFT algorithm is

$$Y[k] = \sum_{i=1}^{N} y[i] e^{-j2\pi ki/N}, \quad 1 \le k \le N \tag{9}$$

where N is the length of the input vector. Here, it is set to be 2048. Thus, the valid amount of data of each frame after FFT processing is 1024, which is still a little big. In order to further reduce the amount, data compression processing should be adopted. FFT spectrum consists of two parts: 1) spectral envelope and 2) spectral details. Considering that the spectral envelope, rather than the spectral details, carries the identification attributes of sound, so it is very important to extract the spectral envelope, which will be used to identify different sounds and reduce data redundancy. PAF is an effective method for envelope extraction and data compression, hence it is adopted in this article. The specific operation method is as follows.

The frequency range of the FFT amplitude–frequency curve is 0–12 kHz, and it is divided into segments every 250 Hz, then all the corresponding amplitudes of each segment are

averaged. Finally, 48 average values can be obtained for each frame, and they are the elements of the envelopes. These 48 average values will form a new envelope vector (***FV***), as shown in (10), so that the data of each frame is compressed by about 97.66%. Furthermore, if it is divided into segments every 1000 Hz, likewise, 12 average values can be obtained for each frame, and they will form another smaller envelope vector (***FV1***), as shown in (11), so the data compression rate is about 99.41%

$$\mathrm{FV}(i) = \mathrm{mean}(y(\mathrm{floor}(21.33 \times (i-1)+1) : \mathrm{floor}(21.33 \times i))) \quad 1 \le i \le 48, \quad 21.33 = 1024 \times 250 \div 12000 \tag{10}$$

$$\mathrm{FV1}(i) = \mathrm{mean}(y(\mathrm{floor}(85.33 \times (i-1)+1) : \mathrm{floor}(85.33 \times i))) \quad 1 \le i \le 12, \quad 85.33 = 1024 \times 1000 \div 12000 \tag{11}$$

where floor() is the function of downward rounding. The size of the original FFT spectrum vector of each frame is 2048, and 352 continuous spectrum vectors will form a spectrogram, whose size is as big as 2048 * 352. However, after the dimension reduction with PAF processing, a simplified spectrogram with the size of 48 * 352 or 12 * 352 can be obtained. Now that ***IDP*** has been confirmed by the time-domain analysis, the corresponding envelope vectors in the simplified spectrogram with the size of 48 * 352 are adopted as the feature vectors (***ES***), which will be used for the following similarity analysis. Thus, the size of bivector ***ES*** is 48 * *cnt*. In addition, the bivector of the spectrogram with the size of 12 * 352 is named as ***BFV1***, which will be used for the following CNN analysis.

3) Eigenvectors of MFCCs Processing: MFCCs processing is another effective feature extraction method, especially for sound recognition. The Mel filter bank is to simulate the human ear's nonlinear perception of sound. Human hearing is selective to frequency and has different sensitivity to different frequency signals. These filters are not uniformly distributed on the frequency axis. There are more filters with higher resolution in the low-frequency region, but fewer filters with lower resolution in the high-frequency region. Here, the relationship $\boldsymbol{G}(f)$ between the frequency (f) and Mel-frequency (f_{mel}) is

$$f_{\mathrm{mel}} = G(f) = 2595 \times \log_{10}(1 + f/700). \tag{12}$$

The Mel filter bank is composed of several Mel filters, whose shape is triangular. The number (M) of Mel filters is often set to be 22–40, in this article, M is set to be 40. The frequency-amplitude response function $H_m(f_{\mathrm{mel}})$ is defined as

$$H_m(f_{\mathrm{mel}}) = \begin{cases} 0, f_{\mathrm{mel}} < F_{m-1}, 1 \le m \le M \\ \frac{f_{\mathrm{mel}} - F_{m-1}}{F_m - F_{m-1}}, F_{m-1} \le f_{\mathrm{mel}} \le F_m \\ \frac{F_{m+1} - f_{\mathrm{mel}}}{F_{m+1} - F_m}, F_m \le f_{\mathrm{mel}} \le F_{m+1} \\ 0, f_{\mathrm{mel}} > F_{m+1} \end{cases} \tag{13}$$

where $[F_0, F_1, \ldots, F_{M+1}]$ is a linear Mel-frequency vector, F_0 equals 0 and F_{M+1} equals $\boldsymbol{G}(f_s/2)$. According to (9), the corresponding frequency f of the kth data is equal to $k*f_s/N$. Combined with (12), we can get

$$f_{\mathrm{mel}} = 2595 \times \log_{10}\left(1 + \frac{f_s k}{700N}\right). \tag{14}$$

Hence, $H_m(k)$ can be easily obtained based on (13) and (14), and the logarithm $S[m]$ of the power-spectrum output filtered by the mth Mel filter is derived as

$$S[m] = 20 \log_{10}\left(\sum_{k=1}^{N/2} \frac{|Y[k]|^2}{N} H_m(k)\right), \quad 1 \le m \le M. \tag{15}$$

In order to remove the coefficient correlation and separate the envelope and details information of the FFT spectrum, discrete cosine transform (DCT) is adopted, and the formula is defined as

$$D[n] = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} S[m] \cos\frac{\pi n(m-0.5)}{M}, \quad 1 \le n \le M \tag{16}$$

where $\boldsymbol{D}$ is the MFCCs vector. Considering that the key information of the FFT spectral envelope is extracted to the low-frequency coefficients, so the second to the 13th coefficients of $\boldsymbol{D}$ are picked out to constitute the envelope vector (***FV2***) of the frame. The size of ***FV2*** is 12, achieving a data compression rate of 99.41%. Hence, after the dimension reduction with MFCCs processing, another simplified spectrogram with the size of 12 * 352 can be obtained, and the corresponding bivector is named as ***BFV2***.

4) Improved Eigenvectors for CNN Analysis: The duration of each frame is about 85.3 ms, thereby, the time of 42 continuous frames is about 3.58 s, which is long enough to record a single snore signal. Considering that the input data size of the CNN should be not only constant, but as small as possible to reduce the complexity of edge computing, hence, the input data size of the CNN will be set to be 12 * 42.

In order to recognize a single snore signal accurately, a simple CNN can be adopted, and the data partitioning is very essential. That is, the input data with the size of 12 * 42 should be created first. As mentioned above, the starting index *idl* and the ending index *idr* of a single sound signal can be confirmed by the short-time energy analysis, hence, their average *ida* can be calculated. Given that the eigenvectors of PAF and MFCCs processing are both effective, then the improved eigenvector (***EV***) for CNN analysis can be derived as

$$\mathrm{EV} = \omega_1 \times \mathrm{BFV1}[1:12, ida-21 : ida+20] + \omega_2 \times \mathrm{BFV2}[1:12, ida-21 : ida+20] \tag{17}$$

where ω_1 and ω_2 are the weight factors. Here, ($ida - 21$) should be larger than 0 and ($ida + 20$) should be less than 353; otherwise, the data beyond the boundary will be filled with 0. Therefore, after the data partitioning and fusion processing, a series of improved eigenvectors are prepared, which will be used for further CNN analysis. So far, all the eigenvectors have been obtained based on the time–frequency analysis, which will be applied for the following snore recognition.

D. Similarity and CNN Analyses

1) Similarity Analysis: In order to reduce the amount of calculation for CNN analysis, similarity analysis will be performed first. The similarities between two adjacent snore signals will be larger than those between two adjacent non-snoring signals. It means that if the similarities between two adjacent eigenvectors in ***ES*** are all very low, then there is no

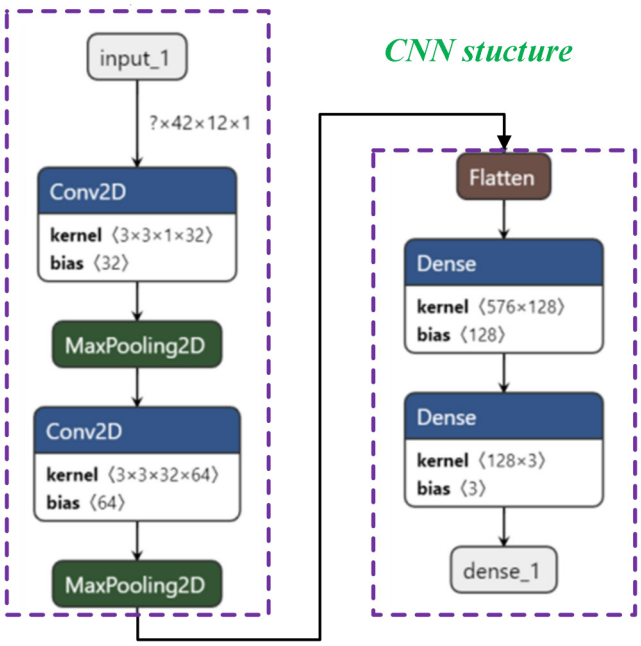


Fig. 8. Lightweight structure of the proposed CNN.

need to conduct further analysis. Compared with CNN analysis, the computation of similarity analysis is simpler, so the preliminary screening with similarity analysis is significant. The formulas of cosine similarity are defined as follows:

$$\begin{cases} C = \cos\theta = \dfrac{\sum_{i=1}^{48} \mathrm{ES}[i][j]\mathrm{ES}[i][j+1]}{\sqrt{\sum_{i=1}^{48} \mathrm{ES}^2[i][j] \sum_{i=1}^{48} \mathrm{ES}^2[i][j+1]}}, 1 \le j < cnt \\ R = \dfrac{\min\left(\sqrt{\sum_{i=1}^{48} \mathrm{ES}^2[i][j]}, \sqrt{\sum_{i=1}^{48} \mathrm{ES}^2[i][j+1]}\right)}{\max\left(\sqrt{\sum_{i=1}^{48} \mathrm{ES}^2[i][j]}, \sqrt{\sum_{i=1}^{48} \mathrm{ES}^2[i][j+1]}\right)}, 1 \le j < cnt \end{cases} \tag{18}$$

where C and R stand for the direction similarity and amplitude similarity between two adjacent eigenvectors, respectively. If they meet the condition (19), the similarities between two adjacent eigenvectors are high, vice versa

$$\begin{aligned} &((\mathrm{RC} > 0.63)\ \&\&\ (C > 0.86)) \\ &\text{or}\ ((C > 0.9)\ \&\&\ (R > 0.65)). \end{aligned} \tag{19}$$

The number of high similarity in a simplified spectrogram with the size of 48 * 352 is recorded as *ND*, if it is not less than 3, it is judged that snoring events may occur in this period and further analysis is necessary; otherwise, there is no snoring event in this period and the snore recognition is over.

2) CNN Analysis and Classification: If the similarities are high enough, then CNN analysis will be further conducted to judge whether there are snore events, and the improved eigenvector ***EV*** with the size of 12 * 42 will be used as the input of the CNN. In order to advance the recognition accuracy and implement the algorithm in the low-cost MCU, a lightweight CNN structure is presented, which is shown in Fig. 8. It is composed of an input layer, two convolution layers, two max-pooling layers, two full-connection layers, a flatten layer, and an output layer. Stride is set to be 1, and no padding processing is adopted. The kernel size is set to be 3 * 3, while the pool size is set to be 2 * 2. Based on Fig. 8, the shape of every layer can be confirmed. The shapes of the input layer, the first convolution layer, and the first max-pooling layer are 42 * 12 * 1, 40 * 10 * 32, and 20 * 5 * 32, respectively. The shapes of the second convolution layer, the second max-pooling layer, and flatten layer are 18 * 3 * 64, 9 * 1 * 64, and 576, respectively. The shapes of the final two full-connection layers are 128 and 3, respectively. That is, the improved eigenvector ***EV*** is the input, and the output will be divided into three categories: 1) snore; 2) noise (other sounds); and 3) quiet.

TABLE I
SAMPLE SIZES OF THE THREE SETS FOR CNN PROCESSING

Type	Sample size for training	Sample size for validation
Snore set	10276	2568
Quiet set	12693	3173
Noise set	9024	2255
Amount	31993	7996

Before classification, the CNN should be trained first. Hence, the sound data sets of snore, noise, and quiet should be applied to the CNN training and validation. As for data acquisition of snore, five mild snorers, five moderate snorers, and five severe snorers were recruited, respectively. In order to imitate the real application scenario, their snores were acquired by the proposed flexible patches depicted in Fig. 1 at home and the experimental tests lasted for about a month. Afterward, the collected data were processed by the methods mentioned above to obtain the improved eigenvectors. Simultaneously, these eigenvectors were labeled manually. Finally, the raw data and data set of snore were formed and stored.

As for data acquisition of noise (other sounds), ten volunteers without snoring were recruited. Similarly, their sleeping noises were acquired by the proposed flexible patches at home and the tests also lasted for about a month. Besides, in order to imitate the real application scenario, the noises of pets, footsteps, talks, music, traffic, and others were also collected. Likewise, all these data were processed to obtain the improved eigenvectors, and they were also labeled manually. Finally, the raw data and data set of noise were formed and stored.

As for data acquisition of quiet, the sounds of the quiet surroundings were collected. Similarly, the raw data and data set of quiet were formed and stored. Finally, the raw data and three data sets are obtained. All these data are anonymized and desensitized (i.e., the sensitive personal information is deleted). The sample sizes of the three sets are shown in Table I, and the sample ratio between the training set and validation set is about 8 to 2. They will be used for a series of experimental analyses.

The batch_size and maximum epoch are set to be 64 and 30, respectively. The learning rate is set to be 1e-4, which remains unchanged in the first ten training epochs and then decreases gradually in the following epochs. The hyperparameters are tuned by experiences to obtain a good training result.

In order to implement the CNN in the ARM MCU, the model parameters need to be simplified. Here, based on Tensorflow Quantization Tool (CubeMX AI), the float-type

TABLE II
SIZES OF FLASH AND RAM REQUIRED BY THE MODEL WITH AND WITHOUT QUANTIZATION

Type	Required Flash (kB)	Required RAM (kB)
Without quantization	363.51	16.98
With quantization	91.54	12.61

parameters in the model are quantized to int8-type to reduce the memory consumption. The sizes of flash and Ram required by the model with and without quantization are shown in Table II. It figures out that after quantization, the resources are greatly reduced. The required flash is decreased to 91.54 from 363.51 kB, while the required RAM is decreased to 12.61 from 16.98 kB. So, the CNN is simple and easy to be implemented in the MCU.

Considering that there are three categories in the output layer, ROC curve, and AUC will not be utilized to evaluate the performances of the network. Instead, the recognition accuracy (ACC) is adopted, which is defined as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{20}$$

where true positive (TP) is the number of correctly detected abnormal epochs (i.e., snore), true negative (TN) is the number of correctly detected normal epochs (i.e., noise and quiet), false positive (FP) is the number of incorrectly detected abnormal epochs (i.e., snore), and false negative (FN) is the number of incorrectly detected normal epochs (i.e., noise and quiet). ACC is a common evaluation index.

E. Snoring Suppression Method

When the body is still, the direction of the head and sleeping posture (such as supine or latericumbent) can be estimated by the components of gravitational acceleration measured by a three-axis MEMS accelerometer. The sampling frequency of the accelerometer is set to be 100 Hz, although it is different from that of the microphone, their signals can be synchronized through the timer interrupt handling of the MCU. If the absolute value of the z-axis acceleration is larger than 0.5 g, the sleeping posture is judged to be supine; otherwise, it is latericumbent. Once snoring events are detected for several times, vibration motor will exert a little force to the head for suppressing snoring, and the anti-snore strategy is as follows.

1) If it is supine, the vibration amplitude is relatively large and the duration is relatively long, since the collapse of the upper airway with supine posture may be severe.
2) If it is latericumbent, the vibration amplitude is relatively small and the duration is relatively short, because the collapse of the upper airway with latericumbent posture may be not severe.
3) Once there are some responses (e.g., large body movement indicated by the accelerometer or stop snoring) or the vibration time is long enough, the interference will stop.

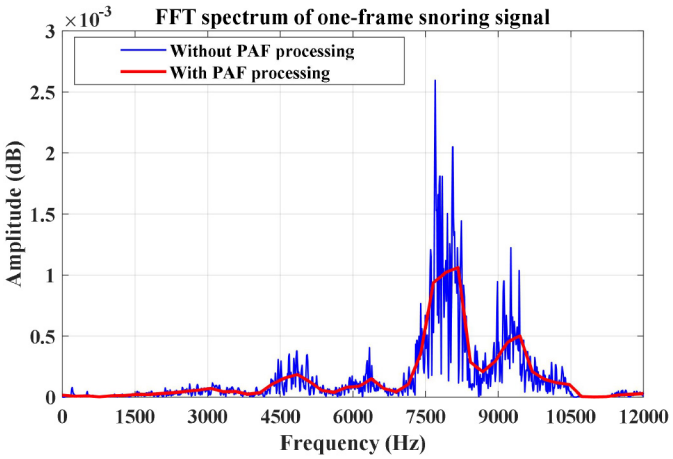


Fig. 9. FFT spectrums of one-frame snoring signal with and without PAF processing.

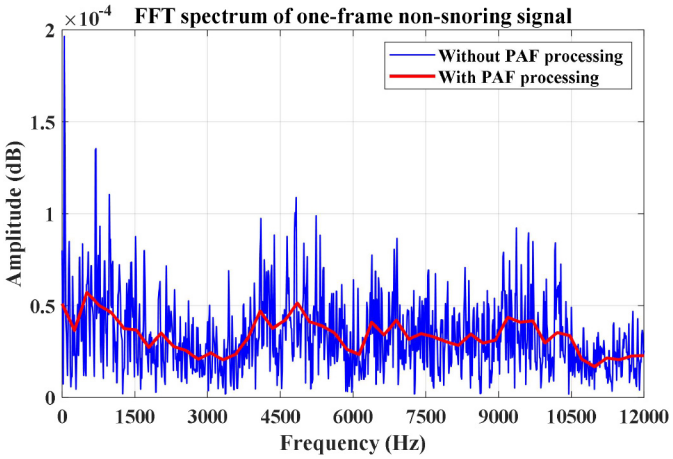


Fig. 10. FFT spectrums of one-frame non-snoring signal with and without PAF processing.

It must be noted that the effect of snoring intervention varies from person to person. For patients with severe snoring, the effect may be general, or even it is ineffective.

IV. EXPERIMENTAL RESULTS

A. Time–Frequency Analysis Results

Based on the acquired experimental data mentioned above, time–frequency analysis can be conducted.

The FFT spectrums of one-frame snoring signal with and without PAF processing are shown in Fig. 9. Besides, the FFT spectrums of one-frame non-snoring signal with and without PAF processing are shown in Fig. 10. In these two diagrams, the blue curve represents the original FFT spectrum vector with the size of 2048, while the red curve represents the envelope vector with the size of 48. They figure out that the main features of the original curve can be extracted effectively by PAF processing, and this dimension-reduction method is feasible.

In addition, the energy distribution of snoring signal is different from that of non-snoring signal. The main energies of the snoring signal are distributed in the frequency band from

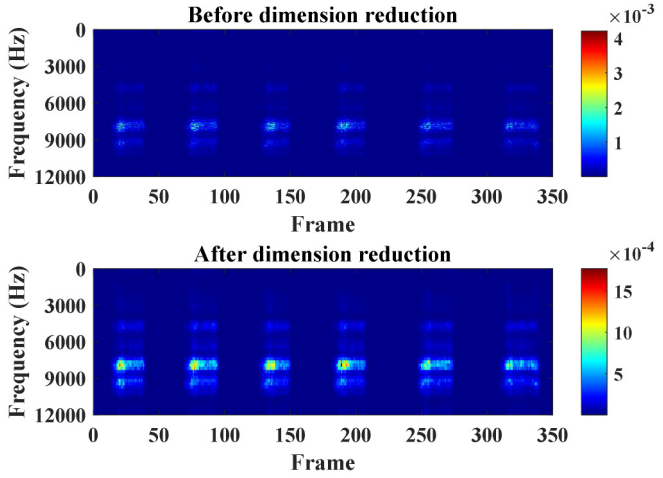


Fig. 11. Spectrograms of the snoring signal with and without PAF processing within 30 s.

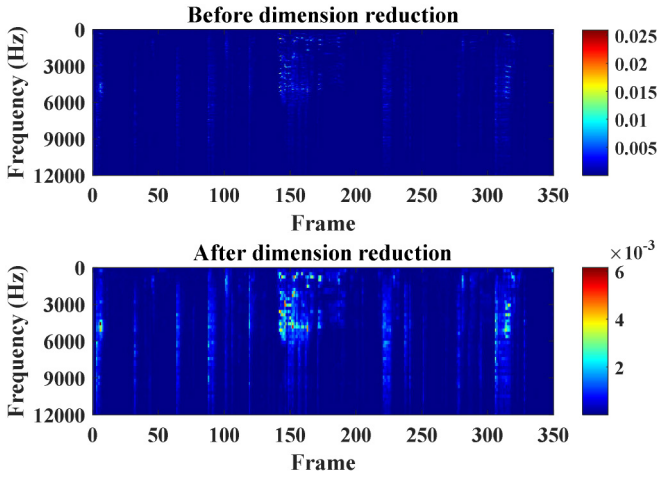


Fig. 12. Spectrograms of the non-snoring signal with and without PAF processing within 30 s.

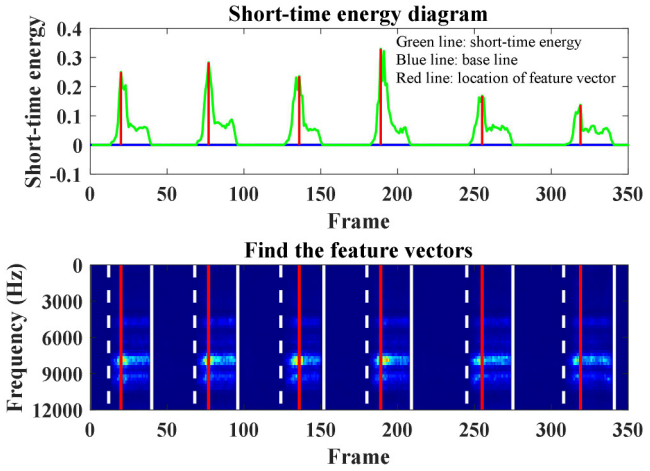


Fig. 13. Short-time energy diagram and the corresponding spectrogram of the snoring signal within 30 s.

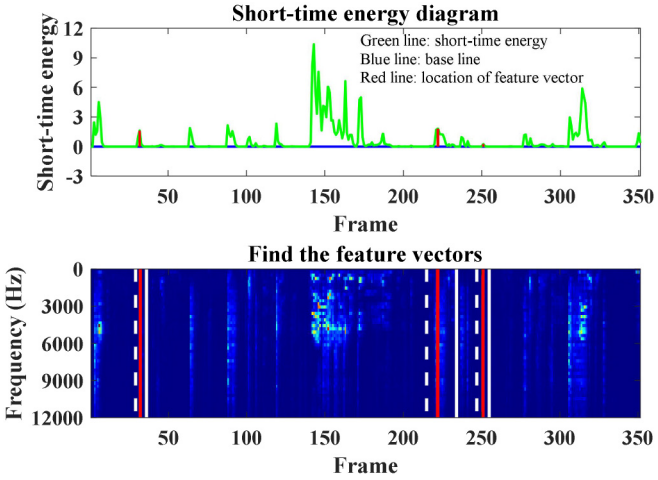


Fig. 14. Short-time energy diagram and the corresponding spectrogram of the non-snoring signal within 30 s.

6 to 10 kHz. However, the energy distribution of the non-snoring signal is not obvious, which is related to the type of the sound. Hence, the envelope vectors after PAF processing can be used as the input for CNN analysis. Besides, the bivector ***ES*** with the size of 48 * *cnt* will be used for similarity analysis.

Moreover, the spectrograms of the snoring and non-snoring signals within 30 s are illustrated in Figs. 11 and 12, respectively. The sizes of the spectrograms with and without PAF processing are 48 * 352 and 2048 * 352, respectively. These two spectrograms figure out that the features of the spectrogram are augmented after dimension reduction, and the energies of the snoring spectrogram in the high-frequency segment are larger than those in the low-frequency segment obviously, which is very different from that of non-snoring spectrogram. In addition, the periodicity in the spectrogram of snoring signal is also stronger than that of non-snoring signal. Thereby, PAF processing is beneficial to the snore recognition.

The short-time energy diagrams and the corresponding spectrograms of the snoring and non-snoring signals within 30 s are depicted in Figs. 13 and 14, respectively. In the energy diagrams, the green curve represents the short-time energy, while the horizontal blue line and vertical red lines stand for the base line and the locations of the eigenvectors, respectively. They demonstrate that the number of detected valid sound envelopes of the snoring diagram is more than that of the non-snoring diagram, and the periodicity of the snoring signal is clearer than that of non-snoring signal.

On the other hand, in the spectrograms, the vertical white dashed lines and solid lines stand for the starting and the ending positions of the single sound signals, respectively. They are determined by the intersection of the green curve and blue line in the energy diagrams. Likewise, the vertical red lines stand for the locations of the eigenvectors (or energy peaks), which are the same as those marked in the energy diagrams, as depicted in Figs. 13 and 14. Thus, the corresponding frame signals are selected, whose FFT spectrum vectors with PAF processing will be chosen as the eigenvectors for the following similarity analysis.

In addition, after locations recognition and data partitioning, the normalized spectrograms of the snoring and non-snoring signals within 3.58 s after MFCCs processing are depicted in Figs. 15 and 16, respectively. They demonstrate that the

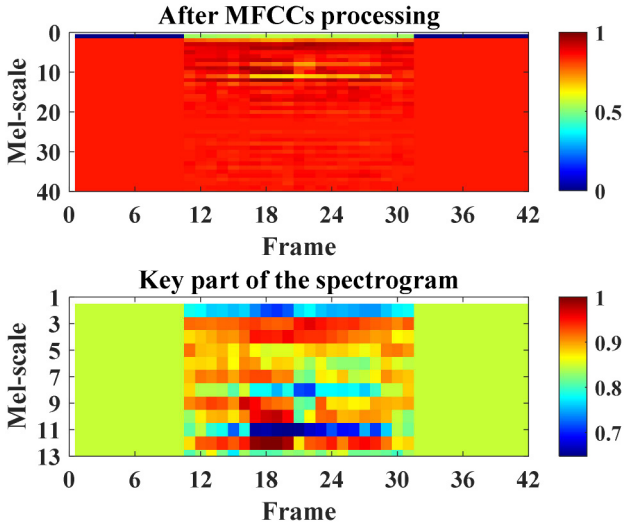


Fig. 15. Spectrograms of the snoring signal within 3.58 s after MFCCs processing.

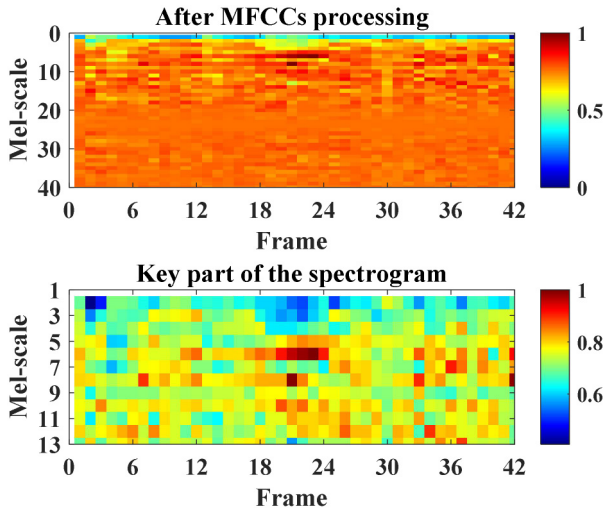


Fig. 16. Spectrograms of the non-snoring signal within 3.58 s after MFCCs processing.

eigenvectors composed of the second to the 13th MFCCs can better represent the features of these frames. Hence, a series of spectrograms with the size of 12 * 42 will be used for further analysis. Taking the eigenvectors of PAF processing into consideration, the improved eigenvectors shown in (17) will be used as the inputs of the CNN to identify whether there are snoring sound envelopes or not.

B. Snore Recognition With Similarity and CNN Analyses

According to Figs. 13 and 14, the similarities between two adjacent eigenvectors in the snoring and non-snoring spectrograms are calculated, as shown in Figs. 17 and 18, respectively. Here, all the *C* and *R* between two adjacent eigenvectors are marked in the spectrograms. If they match the condition (19), they will be dyed purple; otherwise, they will be dyed black. It is clear that the numbers of high similarity in the two spectrograms are 5 and 1, respectively, which indicates that the similarities of the snoring signals are indeed higher than those of the non-snoring signals.

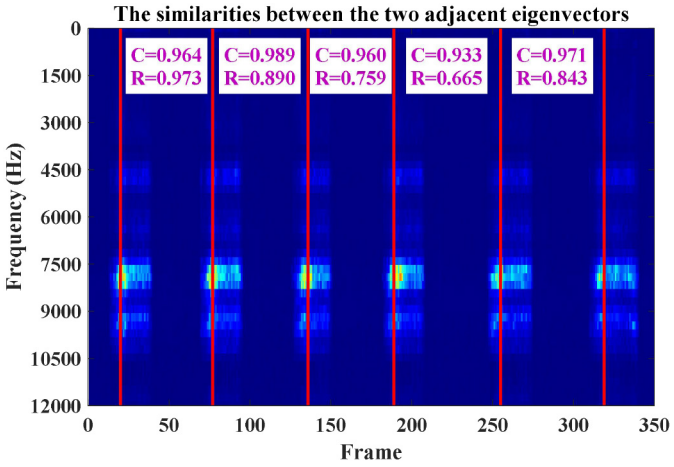


Fig. 17. Similarities between two adjacent eigenvectors in the snoring spectrogram are calculated.

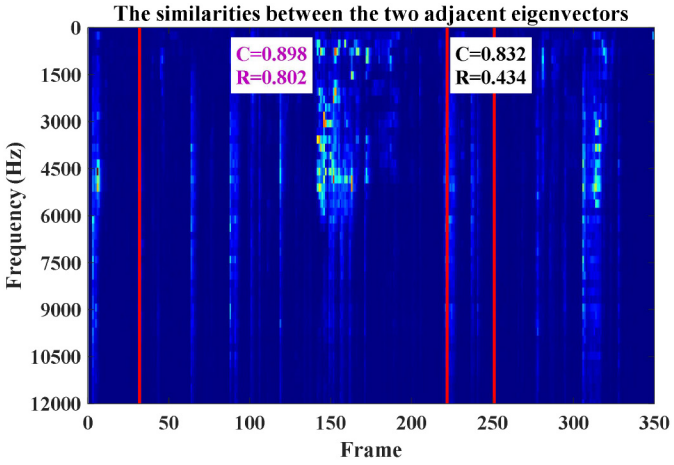


Fig. 18. Similarities between two adjacent eigenvectors in the non-snoring spectrogram are calculated.

The relationship among the direction similarity, amplitude similarity, and the number of high similarity is shown in Fig. 19. Here, 33 samples of the snoring and non-snoring signals within 30 s are selected for analysis. It figures out that *NDs* of the snoring signals are all larger than 3, whereas *NDs* of the non-snoring signals are basically less than 3, only a few of them are not less than 3. Except for the snoring signal, the periodicities or similarities of some other signals may be also obvious or high, that is why a few of *NDs* of the non-snoring signals are larger than 3. Hence, even though *ND is* not less than 3, it is still hard to judge whether there are snoring events or not, and further analysis should be conducted. Therefore, if the number of high similarity *ND* is not less than 3, it is judged that snoring events may occur in this period and CNN analysis will be adopted for further confirmation; otherwise, there is no snoring event in this period and the snore recognition is over. Obviously, the preliminary screening with similarity analysis is feasible. Moreover, it is easy to be realized for the real-time detection since the formulas of similarity analysis are simple.

As for CNN analysis, the accuracy and loss curves of CNN training and validation are depicted in Fig. 20. The sample sizes of the three sets are shown in Table I. The total sample

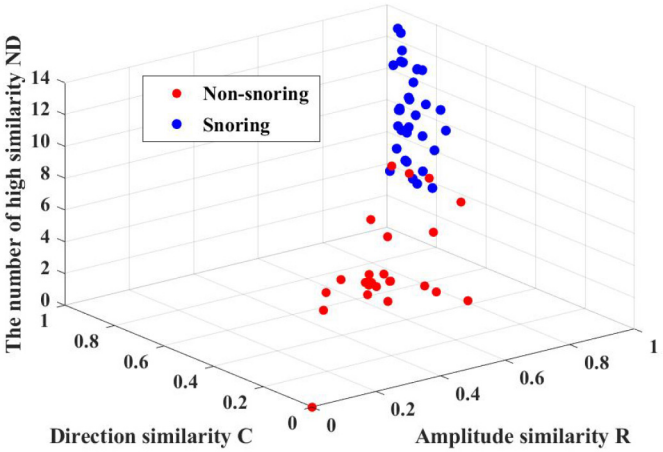


Fig. 19. Relationship among direction similarity, amplitude similarity, and the number of high similarity.

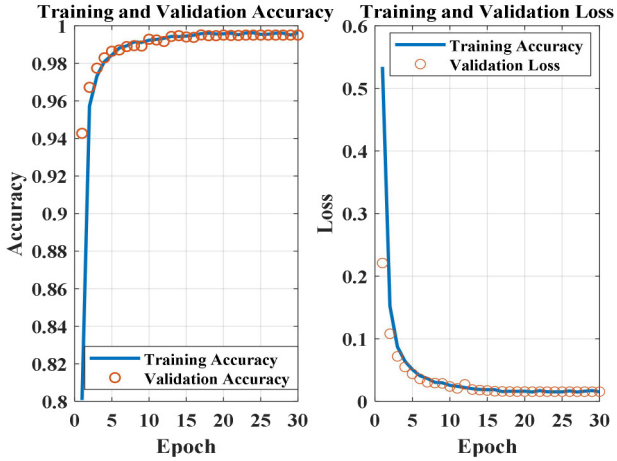


Fig. 20. Accuracy and loss curves of CNN training and validation.

TABLE III
SNORE RECOGNITION RESULTS WITH CNN ANALYSIS

Type	Number of correctly identified	Number of incorrectly identified
Snore validation set	2549	19
Quiet validation set	3160	13
Noise validation set	2227	28
Amount	7936	60

sizes of the training set and validation set are 31993 and 7996, respectively, and the ratio is about 8 to 2. The final accuracies of the training set and validation set are 99.62% and 99.25%, respectively. In addition, the snore recognition results of the validation set with CNN analysis are illustrated in Table III. According to (20) and Table III, ACC is calculated to be as high as 99.25%.

C. Snoring Suppression Result

The snore detection and suppression experiments are performed by a moderate snorer all night long (about 8 h), and the test scenario is shown in Fig. 3. The snoring states of the whole night of the moderate snorer with and without snoring suppression are illustrated in Fig. 21. If he is snoring, the snoring state is marked with 1; otherwise, it is marked with 0. Thus, after three tests, the average snoring times measured by the patches with and without snoring suppression are about 135 and 15 min, respectively, which demonstrates that the snoring suppression method is feasible and effective. Actually, it is very difficult to suppress snoring completely by a single method, and the anti-snoring effect can be further advanced by some combination methods in the future.

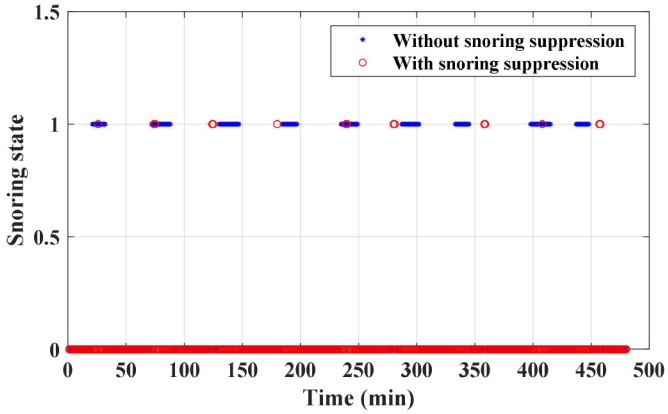


Fig. 21. Snoring states of the whole night of the moderate snorer with and without snoring suppression.

V. CONCLUSION

The contributions of this article include: 1) a novel low-cost flexible patch with MEMS microphone and accelerometer is developed to detect snore event and sleeping posture, and a small vibration motor embedded in the patch is designed to suppress snoring; 2) improved eigenvector, extracted by short-time energy, PAF, and MFCCs processing, is proposed and used as the input of the CNN; and 3) the snore recognition method based on the combination of similarity analysis and CNN analysis is presented, followed by the snoring suppression method. Experimental results demonstrate that the main features of the sound signals can be extracted effectively by PAF and MFCCs processing, and the data compression ratio is about 99.41%. Besides, the locations of the eigenvectors can be found accurately based on short-time energy analysis. The numbers of high similarity of snoring signals within 30 s are larger than 3, while those of non-snoring signals are often less than 3. If the preliminary screening with similarity analysis is passed, CNN analysis will be conducted to judge whether there are snoring events. The accuracy of snore recognition with CNN analysis is calculated to be as high as 99.25%. Finally, the average snoring time measured by the smart patch with snoring suppression is reduced to 15 from 135 min, which indicates that the proposed snore recognition and suppression methods are effective and feasible.

Therefore, this low-cost smart flexible patch can be widely adopted for snoring diagnosis and auxiliary treatment at home. It is wearable, portable, and low cost. Given that the snoring

severity and the effect of snoring intervention vary from person to person, the proposed snore recognition and suppression methods need to be verified and improved repeatedly with a variety of cases of different people. Meanwhile, the CNN model should be continually trained and optimized by more training and validation data sets. In addition, some combination methods for snore recognition should be further developed. For example, in our previous work [56], the respiration rate and amplitude can be measured by a 3-axis MEMS accelerometer, thus auxiliary snore recognition can be performed. However, the difference is that the accelerometer in the sleep monitoring belt is not rotatable, while it is rotatable in the flexible patch. Thus, the detection algorithm should be accordingly improved. The combination of microphone and accelerometer will help to further enhance the snore detection accuracy, and the fusion analysis method is promising. These will be our future work.

REFERENCES

[1] A. Yoshihisa and Y. Takeishi, "Sleep disordered breathing and cardiovascular diseases," *J. Atheroscler. Thromb.*, vol. 26, no. 4, pp. 315–327, Apr. 2019.

[2] J. M. Louis *et al.*, "Predictors of sleep-disordered breathing in pregnancy," *Amer. J. Obstetr. Gynecol.*, vol. 218, no. 5, pp. e1–e12, 2018.

[3] A. Channa, M. Yousuf, and N. Popescu, "Machine learning algorithms for posture identification of obstructive sleep apnea patients using IoT solutions," in *Proc. EHB*, Iasi, Romania, 2020, pp. 1–6.

[4] J. Cambi, Z. M. Chiri, and S. Boccuzzi, "Snoring patterns during home polysomnography. A proposal for a new classification," *Amer. J. Otolaryngol.*, vol. 41, Sep./Oct. 2020, Art. no. 102589.

[5] K. A. Kaplan *et al.*, "When a gold standard isn't so golden: Lack of prediction of subjective sleep quality from sleep polysomnography," *Biol. Psychol.*, vol. 123, pp. 37–46, Feb. 2017.

[6] R. B. Berry, R. Brooks, and C. E. Gamaldo, *AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0*, Amer. Acad. Sleep Med., Darien, IL, USA, 2012.

[7] Z. W. Li, Q. K. Song, L. Cheng, and M. Tan, "Snoring detection based on a stretchable strain sensor," *Sci. China Inf. Sci.*, vol. 64, no. 7, 2021, Art. no. 174201.

[8] E. S. Arnardottir *et al.*, "How to measure snoring? A comparison of the microphone, cannula and piezoelectric sensor," *J. Sleep Res.*, vol. 25, no. 2, pp. 1–11, Apr. 2016.

[9] J. M. Perez-Macias, J. Viik, A. Varri, S. L. Himanen, and M. Tenhunen, "Spectral analysis of snoring events from an Emfit mattress," *Physiol. Meas.*, vol. 37, no. 12, pp. 2130–2143, Nov. 2016.

[10] J. M. Perez-Macias, M. Tenhunen, A. Värri, S.-L. Himanen, and J. Viik, "Detection of snores using source separation on an Emfit signal," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1157–1167, Jul. 2018.

[11] S. Rajala and J. Lekkala, "Film-type sensor materials PVDF and EMFi in measurement of cardiorespiratory signals—A review," *IEEE Sensors J.*, vol. 12, no. 3, pp. 439–446, Mar. 2012.

[12] S. H. Hwang *et al.*, "Polyvinylidene fluoride sensor-based method for unconstrained snoring detection," *Physiol. Meas.*, vol. 36, no. 7, pp. 1399–1414, Jul. 2015.

[13] G. Ye, H. Yin, T. Chen, H. Chen, L. Cui, and X. Zhang, "FENet: A frequency extraction network for obstructive sleep apnea detection," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 8, pp. 2848–2856, Aug. 2021.

[14] A. Smruthy and M. Suchetha, "Real-time classification of healthy and apnea subjects using ECG signals with variational mode decomposition," *IEEE Sensors J.*, vol. 17, no. 10, pp. 3092–3099, May 2017.

[15] X. Wang *et al.*, "Obstructive sleep apnea detection using ECG-sensor with convolutional neural networks," *Multimedia Tools Appl.*, vol. 79, no. 23, pp. 15813–15827, 2020.

[16] C. Wang, J.-H. Chan, S.-H. Fang, H.-T. Cheng, and Y.-L. Hsu, "Novel sleep apnea detection based on UWB artificial intelligence mattress," in *Proc. AICAS*, Hsinchu, Taiwan, 2019, pp. 158–159.

[17] M. Baboli, A. Singh, B. Soll, O. Boric-Lubecke, and V. M. Lubecke, "Wireless sleep apnea detection using continuous wave quadrature Doppler radar," *IEEE Sensors J.*, vol. 20, no. 1, pp. 538–545, Jan. 2020.

[18] M. N. Markandeya, U. R. Aberyratne, and C. Hukins, "Overnight airway obstruction severity prediction centered on acoustic properties of smart phone: Validation with esophageal pressure," *Physiol. Meas.*, vol. 41, no. 10, 2020, Art. no. 105002.

[19] F. Barata *et al.*, "Automatic recognition, segmentation, and sex assignment of nocturnal asthmatic coughs and cough epochs in smartphone audio recordings: Observational field study," *J. Med. Internet Res.*, vol. 22, no. 7, 2020, Art. no. e18082.

[20] M. Młynczak, E. Migacz, M. Migacz, and W. Kukwa, "Detecting breathing and snoring episodes using a wireless tracheal sensor-a feasibility study," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1504–1510, Nov. 2017.

[21] Y. Castillo-Escario, I. Ferrer-Lluis, J. M. Montserrat, and R. Jané, "Entropy analysis of acoustic signals recorded with a smartphone for detecting apneas and hypopneas: A comparison with a commercial system for home sleep apnea diagnosis," *IEEE Access*, vol. 7, pp. 128224–128241, 2019.

[22] S. Z.-E. Z. Ali, R. Ashfaq, R. Afzal, Mehr-UN-Nisa, B. Sultan, and A. Jalil, "Smart pillow: Sleep apnea monitoring & minimization device," in *Proc. RiTA*, Daejeon, South Korea, 2019, pp. 215–219.

[23] S. Vandana, T. Palla, S. Pallempati, and V. Padavala, "Smart pillow," *Int. J. Anal. Exp. Modal Anal.*, vol. 12, no. 8, pp. 2172–2178, Aug. 2020.

[24] Y. Castillo, M. A. Cámara, D. Blanco-Almazán, and R. Jané, "Characterization of microphones for snoring and breathing events analysis in mHealth," in *Proc. EMBC*, Jeju, South Korea, 2017, pp. 1547–1550.

[25] M. A. Camara, Y. Castillo, D. Blanco-Almazán, L. Estrada, and R. Jané, "mHealth tools for monitoring obstructive sleep apnea patients at home: Proof-of-concept," in *Proc. EMBC*, Jeju, South Korea, 2017, pp. 1555–1558.

[26] S. A. H. Tabatabaei, P. Fischer, H. Schneider, U. Koehler, V. Gross, and K. Sohrabi, "Methods for adventitious respiratory sound analyzing applications based on smartphones: A survey," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 98–115, Jun. 2021.

[27] S. Vhaduri, "Nocturnal cough and snore detection using smartphones in presence of multiple background-noises," in *Proc. ACM COMPASS*, Guayaquil, Ecuadaor, 2020, pp. 174–186.

[28] D. Oletic and V. Bilas, "Energy-efficient respiratory sounds sensing for personal mobile asthma monitoring," *IEEE Sensors J.*, vol. 16, no. 23, pp. 8295–8303, Dec. 2016.

[29] U. Saeed, S. Y. Shah, J. Ahmad, M. A. Imran, Q. H. Abbasi, and S. A. Shah, "Machine learning empowered COVID-19 patient monitoring using non-contact sensing: An extensive review," *J. Pharmaceutical Anal.*, vol. 12, pp. 193–204, Apr. 2022. [Online]. Available: https://doi.org/10.1016/j.jpha.2021.12.006

[30] S. Thakur and A. Kumar, "X-ray and CT-scan-based automated detection and classification of covid-19 using convolutional neural networks (CNN)," *Biomed. Signal Process. Control*, vol. 69, Aug. 2021, Art. no. 102920.

[31] M. Gour and S. Jain, "Automated COVID-19 detection from X-ray and CT images with stacked ensemble convolutional neural network," *Biocybern. Biomed. Eng.*, vol. 42, no. 1, pp. 27–41, 2022.

[32] R. A. Abumalloh *et al.*, "Medical image processing and COVID-19: A literature review and bibliometric analysis," *J. Infect. Public Health*, vol. 15, no. 1, pp. 75–93, 2022.

[33] S. Nabavi *et al.*, "Medical imaging and computational image analysis in COVID-19 diagnosis: A review," *Comput. Biol. Med.*, vol. 135, Aug. 2021, Art. no. 104605.

[34] H. Kim, G. Han, and J.-H. Song, "A review for artificial intelligence proving to fight against COVID-19 pandemic and prefatory health policy," *J. Med. Biomed. Appl. Sci.*, vol. 8, no. 8, pp. 494–506, 2020.

[35] K. S. Alqudaihi *et al.*, "Cough sound detection and diagnosis using artificial intelligence techniques: Challenges and opportunities," *IEEE Access*, vol. 9, pp. 102327–102344, 2021.

[36] K. Feng, F. He, J. Steinmann, and I. Demirkiran, "Deep-learning based approach to identify Covid-19," in *Proc. SoutheastCon*, Atlanta, GA, USA, Mar. 2021, pp. 1–4.

[37] Q. Li *et al.*, "MSP-MFCC: Energy-efficient MFCC feature extraction method with mixed-signal processing architecture for wearable speech recognition applications," *IEEE Access*, vol. 8, pp. 48720–48730, 2020.

[38] S. S. Upadhya, A. N. Cheeran, and J. H. Nirmal, "Thomson multitaper MFCC and PLP voice features for early detection of Parkinson disease," *Biomed. Signal Process. Control*, vol. 46, pp. 293–301, Sep. 2018.

[39] S. P. Jin, X. F. Wang, L. L. Du, and D. He, "Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN," *Appl. Acoust.*, vol. 172, Jan. 2021, Art. no. 107562.
[40] Y. E. Erdoğan and A. I. Narin, "COVID-19 detection with traditional and deep features on cough acoustic signals," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104765.
[41] K. K. Lella and A. Pja, "Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath," *Alexandria Eng. J.*, vol. 61, no. 2, pp. 1319–1334, 2022.
[42] T. K. Dash, S. Mishra, G. Panda, and S. C. Satapathy, "Detection of COVID-19 from speech signal using bio-inspired based cepstral features," *Pattern Recognit.*, vol. 117, Sep. 2021, Art. no. 107999.
[43] L. Verde, G. De Pietro, A. Ghoneim, M. Alrashoud, K. N. Al-Mutib, and G. Sannino, "Exploring the use of artificial intelligence techniques to detect the presence of coronavirus Covid-19 through speech and voice analysis," *IEEE Access*, vol. 9, pp. 65750–65757, 2021.
[44] A. Tena, F. Clarià, and F. Solsona, "Automated detection of COVID-19 cough," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103175.
[45] C. Wall, L. Zhang, Y. Yu, and K. Mistry, "Deep recurrent neural networks with attention mechanisms for respiratory anomaly classification," in *Proc. IJCNN*, Shenzhen, China, Jul. 2021, pp. 1–8.
[46] B. Arsenali *et al.*, "Recurrent neural network for classification of snoring and non-snoring sound events," in *Proc. EMBC*, Honolulu, HI, USA, 2018, pp. 328–331.
[47] M. Aly, K. H. Rahouma, and S. M. Ramzy, "Pay attention to the speech: COVID-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings," *Alexandria Eng. J.*, vol. 61, no. 5, pp. 3487–3500, 2022.
[48] J. L. Xie *et al.*, "Audio-based snore detection using deep neural networks," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105917.
[49] H. E. Romero, N. Ma, and G. J. Brown, "Snorer diarisation based on deep neural network embeddings," in *Proc. ICASSP*, Barcelona, Spain, 2020, pp. 876–880.
[50] M. Soltanian and K. Borna, "Covid-19 recognition from cough sounds using lightweight separable-quadratic convolutional network," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103333.
[51] J. Sun, X. Hu, Y. Zhao, S. Sun, C. Chen, and S. Peng, "SnoreNet: Detecting snore events from raw sound recordings," in *Proc. EMBC*, Berlin, Germany, 2019, pp. 4977–4981.
[52] Q. Zhang and R. Boente, "DeepWave: Non-contact acoustic receiver powered by deep learning to detect sleep apnea," in *Proc. BIBE*, Cincinnati, OH, USA, 2020, pp. 723–727.
[53] H. A. Rashid, A. N. Mazumder, U. P. K. Niyogi, and T. Mohsenin, "CoughNet: A flexible low power CNN-LSTM processor for cough sound detection," in *Proc. AICAS*, Washington, DC, USA, Jun. 2021, pp. 1–4.
[54] N. Petrellis and G. K. Adam, "Cough sound classification based on similarity metrics," in *Proc. TSP*, Brno, Czech Republic, Jul. 2021, pp. 214–217.
[55] U. Sait *et al.*, "A deep-learning based multimodal system for Covid-19 diagnosis using breathing sounds and chest X-ray images," *Appl. Soft Comput.*, vol. 109, Sep. 2021, Art. no. 107522.
[56] C. H. He, J. W. Tan, X. L. Jian, G. X. Zhong, L. L. Cheng, and J. Z. Lin, "A smart flexible vital signs and sleep monitoring belt based on MEMS triaxial accelerometer and pressure sensor," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 14126–14136, Aug. 2022.

Chunhua He received the B.S. degree in microelectronics from Sun Yat-sen University, Guangzhou, China, in 2010, and the M.S. degree and the Ph.D. degree in microelectronics from Peking University, Beijing, China, in 2013 and 2018, respectively.

From 2013 to 2017, he was an Engineer with the No. 5 Electronics Research Institute, Ministry of Industry and Information Technology, Guangzhou. From 2017 to 2019, he was an Engineer with Midea Group, Foshan, China. From 2019 to 2021, he was a Senior Engineer with Guangzhou 37 Degree Smart Home Company Ltd., Guangzhou. He joined the School of Computer, Guangdong University of Technology, Guangzhou, in 2021, where he is currently an Associate Professor. His current research interests include the design and application of MEMS sensors and AI algorithms for biomedical engineering.

Jiewen Tan received the B.S. degree in mechanical engineering and the M.S. degree in material processing engineering from Huazhong University of Science and Technology, Wuhan, China, in 2014 and 2016, respectively.

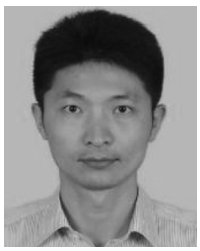
From 2016 to 2019, he was an Engineer with Midea Group, Foshan, China. He joined Guangzhou 37 Degree Smart Home Company Ltd., Guangzhou, China, in 2019, where he is currently an Engineer. His current research interests include intelligent sleep monitoring and biological information processing.

Xuelei Jian received the B.S. degree in mechanical manufacturing and automation from Wuhan Institute of Technology, Wuhan, China, in 2017, and the M.S. degree in mechanical engineering from Shantou University, Shantou, China, in 2020.

He joined Guangzhou 37 Degree Smart Home Company Ltd., Guangzhou, China, in 2020, where he is currently an Engineer. His current research interests include machine structure design and automatic speech recognition, computer vision, and continuous glucose measurement.

Guangxiong Zhong received the B.S. degree in electronic information engineering from Shantou University, Shantou, China, in 2009. He is currently pursuing the M.S. degree in technology economy and management with South China University of Technology, Guangzhou, China.

From 2009 to 2015, he was an Engineer with Southern Power Group Technology Development Company Ltd., Guangzhou. From 2015 to 2017, he was an Engineer with Guangdong Hengju Medical Technology Company Ltd., Guangzhou. From 2017 to 2019, he was an Engineer with Midea Group, Foshan, China. He joined Guangzhou 37 Degree Smart Home Company Ltd., Guangzhou, in 2019, where he is currently a Senior Engineer. His current research interests include embedded intelligent product design, design and application of multisensor fusion algorithms, and artificial intelligence algorithms in biomedical engineering.



Heng Wu received the B.S. degree in electronic and information engineering from Wuhan University of Science and Technology, Wuhan, China, in 2009, and the M.S. and Ph.D. degrees in optics and mechanical manufacture and automation from the South China University of Technology, Guangzhou, China, in 2012 and 2017, respectively.

He joined the School of Automation, Guangdong University of Technology, Guangzhou, in 2017, where he is currently an Associate Professor. His research interests are in the fields of optical imaging, optical measurement, machine vision, and image processing.

Lianglun Cheng received the B.S. degree and the M.S. degree in automatic control engineering from Huazhong University of Science and Technology, Wuhan, China, in 1988 and 1992, respectively, and the Ph.D. degree in automatic control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1999.

He joined the School of Computer, Guangdong University of Technology, Guangzhou, China, in 1992, where he is currently a Professor and the Dean. His current research interests include Internet of Things and information physical integration system, interconnection and fusion of heterogeneous network, big data of industrial process, and high-performance computing.

Prof. Cheng is currently an expert with special subsidy from the Government of the State Council, the Director of the National and Local Joint Engineering Research Center for Integrated Technology of Intelligent Manufacturing Information Physics Fusion System, the Director of the Key Laboratory of Guangdong Information Physics Fusion System, and the Deputy Director of the Qian Xuesen Innovation Committee of CASS, the Director of China Automation Society, a member of China Computer Society, the Vice Director of Guangdong Automation Society, and the Executive Vice President of Guangdong Robotics Society.

Juze Lin received the B.S. degree in health and rehabilitation of traditional Chinese medicine and the M.S. degree and the M.D. degree in clinical specialty of integrated Chinese and Western medicine from Guangzhou University of Chinese Medicine, Guangzhou, China, in 2003, 2006, and 2018, respectively.

He joined Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences and Guangdong Institute of Gerontology, Guangzhou, in 2006, where he is currently a Deputy Chief Physician of Clinical Specialty of Integrated Chinese and Western Medicine with the Department of Traditional Chinese Medicine. His current research interests include the integrated Chinese and Western medicine, and biomedical engineering.

Dr. Lin is currently an Editorial Board Member of *Journal of Chinese Folk Therapy*; a member of the Professional Committee of Cancer Palliative Care Research, World Federation of Chinese Medicine Societies; the Director of Oncology Medical Professional Committee of Chinese Association of Ethnic Medicine; the Vice Secretary General of Geriatrics Committee of Chinese Association of Ethnic Medicine; the Vice Chairman of Geriatric Disease Prevention Committee of Guangdong Traditional Medical Association; and the Vice President of Guangdong Society of Folk Traditional Chinese Medicine.