SymRecorder: Detecting Respiratory Symptoms in Multiple Indoor Environments Using Earphone-Microphones

Zhiyuan Li[1], Feng Hong[1(✉)], Yan Xue[2], Qingbin Li[1], and Zhongwen Guo[1]

[1] Ocean University of China, Qingdao 266000, China
{lizhiyuan,liqingbin}@stu.ouc.edu.cn, {hongfeng,guozw}@ouc.edu.cn
[2] Qingdao University, Qingdao 266100, China
xueyan0512@qdu.edu.cn

Abstract. Respiratory symptoms associated with sound frequently manifest in our daily lives. Despite their potential connection to illness or allergies, these symptoms are often overlooked. Current detection methods either depend on specific sensors that must be deliberately worn by the user or are sensitive to environmental noise, limiting their applicability to specific settings. Considering that indoor environments vary, we propose SymRecorder, an earphone microphone-based application, for detecting respiratory symptoms across a range of indoor settings. By continuously recording audio data through the earphone's built-in microphone, we can detect the four common respiratory symptoms: cough, sneeze, throat-clearing, and sniffle. We have developed a modified ABSE-based method to detect respiratory symptoms in noisy environments and mitigate the impact of noise. Additionally, a Hilbert transform-based method is employed to segment the continuous respiratory symptoms that users may experience. Based on selected acoustic features, the four symptoms are classified using the residual network and the multi-layer perceptron. We have implemented SymRecorder on various Android devices and evaluated its performance in multiple indoor environments. The evaluation results demonstrate SymRecorder's dependable ability to detect and identify users' respiratory symptoms in various indoor environments, achieving an average accuracy of 92.17% and an average precision of 90.04%.

Keywords: respiratory symptoms · ear-phone · signal process · deep-learning

1 Introduction

Respiratory symptoms are associated with illnesses, infections or allergies. For example, cough is the main symptom of asthma. When the patient has pneumonia (e.g., COVID-19), it is often accompanied by throat-clearing (t-c) and

L. Wang et al. (Eds.): CWSN 2023, CCIS 1994, pp. 301–315, 2024.
https://doi.org/10.1007/978-981-97-1010-2_22

nasal aspiration symptoms. Currently, patients commonly use subjective reporting methods when seeking medical care [6]. This has been shown to be inefficient and inaccurate.

In recent years, works have focused on the detection of specific types of respiratory symptoms, such as cough [13], sneeze [1], and snore [15]. PulmoTrack-CC [14] uses a combination of sound recorded from the neck and a motion sensor placed on the chest to achieve a sensitivity of approximately 96% when calculating cough events. All of the above systems require the user to wear special sensors and are not practical enough. With the increasing power of smartphones, many studies have emerged on the use of smartphones to improve the quality of healthcare services [10,11,17,18]. A cough detection system [19] uses a local Hu matrix and a k-nearest neighbor (KNN) algorithm to achieve 88.51% sensitivity (SE) and 99.72% specificity (SP). SymDectector [12] is a smartphone-based application that implements the detection of sound-related respiratory symptoms in office and home scenarios. SymListener [16] implements three types of respiratory symptom detection in driving environments with strong interior noise. However, SymDetector and SymListener do not consider continuous symptoms. The popularity of earphones provides an opportunity to detect respiratory symptoms in multiple indoor environments. When users wear earphones, their relative position to the human body does not change and they are able to receive the acoustic signals generated by the user more stably.

Driven by these circumstances, we propose an earphone-microphone based system, called SymRecorder, for detecting sound-related respiratory symptoms in a variety of indoor environments. SymRecorder uses the earphone microphone connected to a smart device to sense the environment and detects and recognizes sound-related respiratory symptoms, including cough, sneeze, t-c and sniffle. To achieve the above objective, we face the following challenges: (1) the indoor environment where the user is located may be noisy, which can lead to a lower signal-to-noise ratio and make it difficult to detect sound events; (2) the user may experience continuous respiratory symptoms, especially continuous cough, at very short intervals, SymRecorder needs to accurately subdivide these continuous symptoms.

To address the above challenges, we design a sound event detection method combining dual threshold and Adaptive Band-partitioning Spectral Entropy (ABSE) [3], named RA-ABSE to detect sound-related events occurring in different indoor environments. RA-ABSE uses dual thresholds to detect sound event endpoints in quiet environments. While in the noisy environment, ABSE is used as a feature to detect the endpoints of sound-related events and is combined with Berouti power spectrum subtraction to remove the effect of noises on sound events. With the help of the RA-ABSE, segments of the audio containing sound events are filtered out. After acquiring the sound event fragments, we design a Hilbert Transform (HT) based method to subdivide the possible continuous symptoms. Then we use a combination of features based on Mel Frequency Cepstrum Coefficients (MFCC), Gammatone Frequency Cepstrum Coefficients (GFCC) and spectrogram. SymRecorder adopts the Residual Network (ResNet) and Multi-layer Perceptron (MLP) to classify the four types of respiratory symptoms. We also incorporate the attention mechanism into the ResNet to highlight the unique features of the same respiratory symptoms and reduce the influence of different environments and different populations.

To evaluate the performance of SymRecorder, we collect data from a total of 20 volunteers over 4 months using earphones to build the system model. We implement SymRecorder on the Android platform and comprehensively evaluate its performance. The experimental results show that SymRecorder is effective in four indoor environments: home, office, canteen, and shopping mall. Our contributions are summarized as follows:

- We propose a detection system, called SymRecorder to detect sound-related respiratory symptoms in different indoor environments. Through acoustic sensing, SymRecorder only uses a pair of earphones and a mobile device to detect and differentiate between cough, sneeze, t-c, and sniffle.
- We design a dual threshold and ABSE-based sound event detection method, called RA-ABSE, to detect sound events in different indoor environments, and use Berouti power spectrum subtraction to eliminate the environmental noises. We also design a HT based method to subdivide possible continuous respiratory symptoms.
- We design a combination of features based on the spectrogram, MFCC, and GFCC, and use a deep learning model combining ResNet, attention mechanism, and MLP for classification. The evaluation results show that SymRecorder has an average accuracy of 92.17% and an average precision of 90.04%.

The rest of this article is organized as follows: In Sect. 2, we describe The detailed description of the SymRecorder design. Experimental details and future work on SymRecorder are presented in Sect. 3. Section 4 discusses related work, and finally, we draw our conclusion in Sect. 5.

2 System Design

This section describes the system architecture of SymRecorder. As shown in Fig. 1, the whole system consists of six modules. First, the original microphone audio recording is split into frames and windows, and the frames and windows are sent to the sound event detection module. This module first determines the current environment type and detects sound events using the RA-ABSE method. Next, the sound events are passed through the continuous symptom detection module to subdivide the possible continuous symptoms. Next, features are extracted for each filtered sound event and a deep learning network is used to classify the sound events. Finally, respiratory symptoms are recorded. The design details of each module are described in detail below.

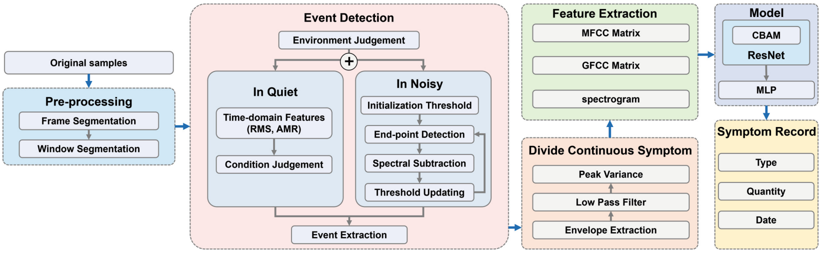
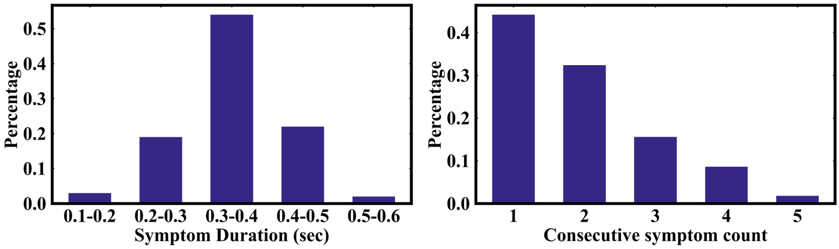


Fig. 1. System overview.

2.1 Sampling and Pre-processing

Existing earphones are capable of sampling audio signals at a variety of sampling rates. We choose 20000 Hz as the sampling rate. The sampled audio stream is then segmented into 10ms non-overlapping frames, which are used to extract time-domain features. The VocalSound [7] dataset contains recordings of 3365 subjects performing six physiological activities: laugh, sigh, cough, t-c, sneeze, and sniffle. We count the distribution of all symptom durations. As seen in Fig. 2a, respiratory symptoms typically last for hundreds of milliseconds and cover multiple frames. Therefore, we group a fixed number of consecutive frames into a single window for processing. In addition, the user may also experience continuous respiratory symptoms, especially continuous cough. To determine the window size, we also count the number of possible occurrences of continuous respiratory symptoms. As shown in Fig. 2b, continuous respiratory symptoms tend to last 1 to 3 times, while reference [5] states that during continuous respiratory symptoms, subsequent symptoms will not include an inspiratory period except for the first symptom, and the duration of each symptom will not exceed 0.5 s. Therefore, the window size is set to 2 s, which can cover any respiratory symptoms. To avoid double counting, there is no overlaph between windows. When a user experiences consecutive symptoms, they are distributed in a maximum of two windows.



(a) the distribution of symptom duration

(b) the distribution of consecutive symptom count

Fig. 2. The distribution of symptom.

2.2 Sound Event Detection

We utilize the short-time power (STE) to determine the user's environment. Specifically, SymRecorder stores data from the current window and the past 4 windows, totaling 5 windows (i.e., 10 s). Subsequently, the STE of the frames within each window is computed. Only when 80% of the frames' STE in each window are below the STE threshold (i.e., 10), the current environment is classified as a quiet environment. Otherwise, it is considered as a noisy environment. Following this, we design a sound event detection method called RA-ABSE. In quiet indoor environments, the method employs dual-threshold time-domain features for sound event detection, while in noisy indoor environments, ABSE is used as a feature to detect sound events.

Quiet Indoor Environment. In a quiet indoor environment, the energy of the audio signal received by the earphone microphone is typically low except for sound events. Figure 3a illustrates an earphone audio recording in an office scenario with a subject's speech signal and several respiratory symptoms. It can be observed that the energy of the environmental noise is very low except for the sound events. Furthermore, in addition to discrete sound events containing respiratory symptoms, continuous sound events (e.g., speech or music) are included, which need to be filtered out. In the following, we introduce the employed time-domain features and elucidate how these features can be used to filter out continuous sound events.

Root Mean Square (RMS): Suppose l denotes the frame consisting of N samples, and $x(l, n)$ denotes the amplitude value of the n sample in l, then the RMS [8] of the l frame is

$$rms\left(l\right) = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x\left(l, n\right)^2} \tag{1}$$

The RMS measures the energy level contained in the current acoustic frame so that the RMS can distinguish between acoustic and non-acoustic event frames.

Above α-Mean Ratio (AMR): Assuming that w represents a window consisting of m frames, the AMR of the window w is calculated as

$$amr\left(\alpha, w\right) = \frac{\sum_{i=1}^{m} ind\left[rms\left(l_i\right) > \alpha \cdot \overline{rms}\left(w\right)\right]}{m} \tag{2}$$

where $\overline{rms}\left(w\right)$ is the mean RMS of all frames in window w and $ind\left(\cdot\right)$ indicates the indicator function that returns 1 when the condition is true and 0 otherwise. α is the given parameter. AMR measures the ratio of high-energy frames in the window and the experimental parameter α is jointly set with the mean RMS of the window to distinguish between high-energy and low-energy windows. Given an appropriate value of α, windows containing discrete sound events, continuous sound events, and environmental noise return different AMR. Therefore, this feature can be used to filter windows with discrete sound events. In SymRecorder, α is set to 0.6.

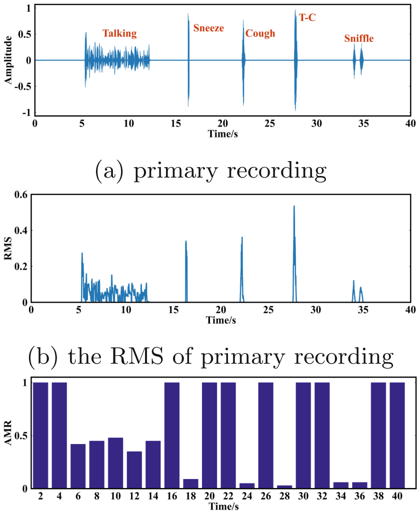
RMS is first used to find the endpoints of sound events. As shown in Fig. 3b, sound events usually have higher energy, and therefore the RMS of the sound event frames is also significantly larger than the surroundings. Specifically, when the RMS of three consecutive frames is above the RMS threshold β (i.e., 0.005), the beginning of the first frame is considered the start point of the sound event. The end point is obtained when the RMS of three consecutive frames below the threshold. And the AMR is used to filter out continuous sound events, especially the user's speech signals. As shown in Fig. 3c, windows contain discrete sound events typically have lower AMR due to the windows contain fewer frames of sound events, while the sound events contain much more energy than environmental noise frames. The AMR of the speech event window typically ranges from 0.3 to 0.5, since the voiced frames occupy about 30% to 50% [9] in a fluent speech. Therefore, when the AMR of the window where the current sound event is less than 0.3 and the duration of the sound event is greater than 0.2 s, the sound event is considered as a valid sound event, otherwise, the sound event is discarded.

Finally, we consider the situation when the user experiences continuous symptoms. We observe that when most of the continuous symptoms are distributed across two windows, the AMR of the window containing more symptom parts will be slightly higher, but still below the threshold of 0.3, so that the continuous symptoms are preserved. However, when continuous symptoms are concentrated within a single window, the AMR of that window becomes similar to the AMR of the continuous speech windows, which means that the continuous symptoms will be discarded. Therefore, if the AMR of the window containing the sound event is higher than 0.3 but the duration of that sound event is lower than the window size (i.e., 2 s), the sound event is still preserved.

Noisy Indoor Environment. In a noisy indoor environment, the earphone microphone continuously receives audio signals with higher energy. Figure 4a shows an audio recording in a canteen scene, where it can be observed that the environmental noise in the canteen makes it challenging to accurately detect respiratory symptoms by time-domain features. Therefore, we employ the ABSE as a feature parameter to detect sound events in noisy environments. ABSE divides the spectrum into multiple frequency bands and calculates the spectral entropy within each frequency band, thus avoiding dependence on the entire spectral amplitude variance. The ABSE for the l frame is calculated as

$$H_b(l) = \sum_{m=1}^{N_b} W(m,l) \cdot H_b(m,l) \tag{3}$$

where $W(m,l)$ and $H_b(m,l)$ are the weight and spectral entropy value of the m sub-band, respectively. Then an adaptive signal threshold T_s is set to classify

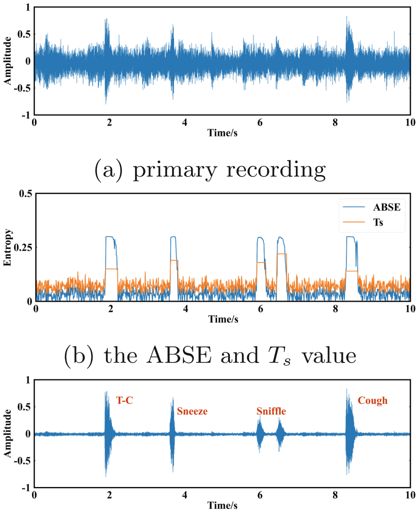


(a) primary recording

(b) the RMS of primary recording

(c) the AMR of primary recording

Fig. 3. Example of audio recording in office.



(a) primary recording

(b) the ABSE and T_s value

(c) spectrum after noise cancellation

Fig. 4. Example of audio recording in canteen.

event segment or noise-only segment according to the mean μ and variance θ of the logarithmic ABSE value of detected noise-only segments. Formally, $T_s = \mu + \gamma \cdot \sigma$, and γ (i.e., 0.005) is an experimental coefficient. This threshold is compared to the value of the current frame. Whenever the difference surpasses a specified threshold, event segment is detected. If a given segment is detected as a noise-only segment, then the signal threshold is updated. Figure 4b illustrates the trend of the ABSE of Fig. 5(a) and T_s. It can be observed that T_s is constantly updated during the pure noise and remains unchanged when a sound event is detected.

After the sound events endpoints are detected, it is necessary to separate the noise components from the sound events. We employ the Berouti spectral subtraction method to reduce the noise components in the sound events. Suppose $Y(e^{jw})$, $S(e^{jw})$ and $N(e^{jw})$ denote the Fourier Transform (FT) result of the mixed noisy signal, the pure signal and the additive noise, then we have $|Y\left(e^{jw}\right)|^2 = |X\left(e^{jw}\right)|^2 + |N\left(e^{jw}\right)|^2$. As for the additive noise can not be obtained directly, we use the average power spectral E of several beginning frames to approximately replace $|N\left(e^{jw}\right)|^2$. Finally, $|S(e^{jw})|$ can be calculated by $|S(e^{jw})| = \sqrt{|Y\left(e^{jw}\right)|^2 - E}$. Figure 4c illustrates the processed result, it can be seen that most of the environmental noise has been eliminated, and the sound events can be effectively extracted from the time domain.

2.3 Subdivision of Continuous Symptom

Although continuous symptoms mainly refer to continuous cough, in order to cope with other continuous symptoms that may occur, all detected sound events are sent to this module. We design an algorithm based on HT to subdivide the sound events that may contain continuous respiratory symptoms.

The algorithm execution steps are shown in Fig. 5. Firstly, the HT is applied to the sound events detected in the previous stage to extract the envelope, representing the amplitude contour of the sound events. The HT is applied to smooth the sound signal and eliminate the negative values [4].

The envelope is then passed through a Butterworth low-pass filter, as a way to obtain the fundamental frequencies of the continuous respiratory symptoms. The frequency range of the low-pass filter is estimated based on the duration of the current sound event. Assuming the duration of the current sound event is t, as shown in Fig. 2b, the number of possible occurrences of consecutive symptoms is from 1 to 4. Thus, the frequency interval of the current symptom during the time of t is (1/t, 4/t) Hz. We set this frequency interval as the frequency range of the low-pass filter and iteratively increment 0.1 Hz. When the filter frequency approaches the frequency of the current respiratory symptoms, the number of peaks on the filtered envelope corresponds to the number of occurrence counts of the current symptom. Thus, when the criteria for the number of peaks are met, the variance of all peaks is recorded until the iteration process concludes. The set of peaks with the minimum variance is subsequently selected, and the subdivision of sound events is achieved by the distance differences between the peaks. In the algorithm design process, additional conditional statements are incorporated to handle specific situations:

(a) Since the number of peaks in the filter envelope corresponds to the number of times during the filter frequency change, only one peak can be detected for a sound event that contains only one respiratory symptom. If only one peak is still detected when the filter frequency iterates to 4/tHz, the current sound event is not processed in the current module.

(b) Some single symptoms can have two stages of energy bursts, with the first phase being sharper and containing higher energy, while the second stage is relatively flat and has lower energy. Therefore, two peaks may appear during the filter frequency iterations, indicating the subdivision of a single symptom. To differentiate it from two consecutive symptoms, the values of the two peaks are compared after the set of peaks with the lowest variance is obtained. For two consecutive symptoms, the peaks on the filtered envelope will be evenly distributed. Suppose the first peak value is $Peak_1$ and the second peak value is $Peak_2$. If $0.8 \cdot Peak_1 < Peak_2$ is satisfied, the sound event is subdivided according to the distance between the peaks, otherwise the current sound event is output directly.

(c) Two stages of energy bursts may also occur during continuous symptoms. The variance of the peak set can help filter out such case. During the iteration of the filter frequency, when a smaller peak appears, the variance of the set of peaks increases, and therefore the current peak set is not selected. So, if the number of selected peak set is more than two, the current sound event is subdivided directly according to the distance between the peaks.

Finally, we perform alignment processing on each subdivided sound event to facilitate the next step of feature extraction. Specifically, the duration of the sound event is denoted as d, if $d < 0.2\,\mathrm{s}$, the sound event is discarded; if $0.2\,\mathrm{s} < d < 0.5\,\mathrm{s}$, then the sound event is zero-padded to $0.5\,\mathrm{s}$; if $d > 0.5\,\mathrm{s}$, then the middle part of the sound event is taken, and the part before and after the length of $1/2 \cdot (d - 0.5)$ is deleted.

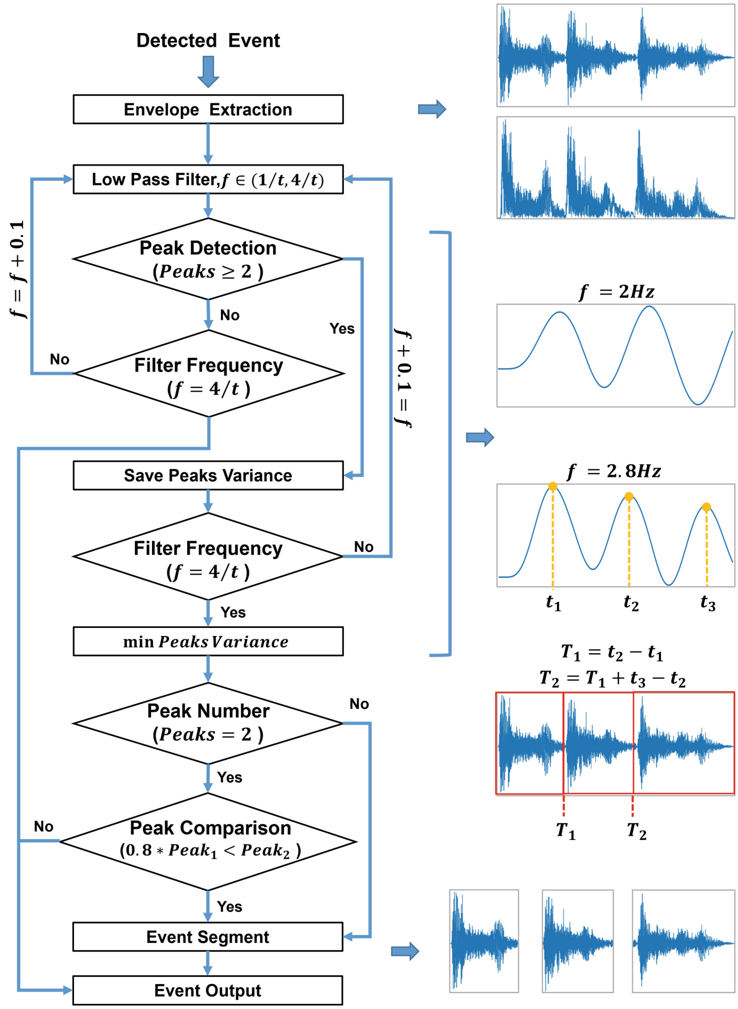


Fig. 5. Subdivision algorithm.

2.4 Feature Extraction and Classified Model

Respiratory symptoms are abnormal manifestations related to the respiratory system, typically emitted from the nasal cavity or throat, presented in the acoustic form of specific audio signals. Many features exist for identifying specific types of audio signals, and one of the most commonly used features is the MFCC. MFCC takes into account the non-linear response of the human ear on the audio spectrum, and is obtained through a frequency transformation of the logarithmic spectrum.

Although MFCC is widely used as a feature in audio signal processing, the performance of MFCC is strongly influenced by the noise level. The Gammatone filter bank can provide higher accuracy compared to the Mel filter bank. To make the acoustic features more robust, we also use GFCC as a feature.

In addition, SymRecorder requires a feature to describe the local information of respiratory symptoms in both frequency and time domains. Short-term Fourier Transform (STFT) splits the original signal into fixed-length time windows and applies the FT, which can capture the short-time spectral features in the original signal.

SymRecorder uses deep learning networks to capture the distinctive representations of each respiratory symptom. The network architecture is shown in Fig. 6. The learning network uses Convolutional Neural Network (CNN) and ResNet as the backbone, MFCC matrix, GFCC matrix, and spectrogram as inputs. To enhance the differences between different sound event features, the lightweight Convolutional Block Attention Module (CBAM) is integrated into the ResNet. Finally, the fine-grained features extracted by the learning network are concatenated into the same feature vector and then classified using the MLP.

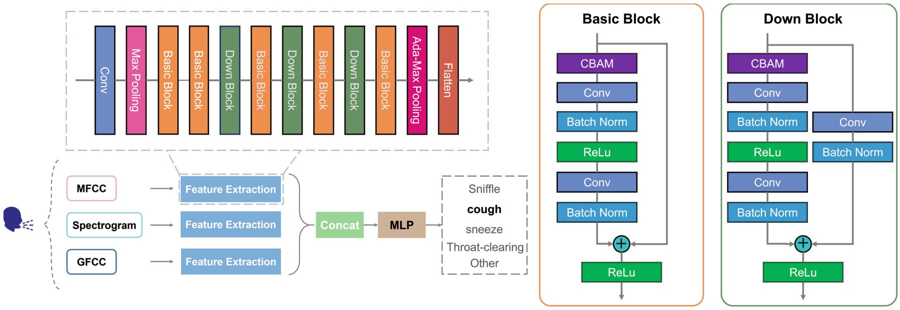


Fig. 6. Finer feature extraction and symptom identification network structure.

3 Experimentation and Evaluation

In this section, we present the implementation details and evaluates the performance of SymRecorder based on the data collected from experiments. We also conclude with a discussion on the future work of SymRecorder.

3.1 Experimental Setup

The training set used is derived from two datasets. The first dataset is from VocalSound [7]. We get 2013 cough, 1310 sneeze, 1764 t-c, and 2341 sniffle samples. These samples are utilized to investigate the features extracted from respiratory symptoms and enable deep learning networks to learn the distinctions between different respiratory symptom characteristics.

The second dataset comes from 14 participants we recruited, consisting of 4 females and 10 males. The participants' ages range from 12 to 58 years. Three participants are from the same family and spend much of their time at home; the remaining 11 participants are graduate students who frequented the office and canteen almost every day. Additionally, they spend one day per week shopping at the mall. Over a period of four months, we collect 2,873 cough, 2,008 sneeze, 2,577 t-c, and 3,135 sniffle samples under the four different environmental conditions. In addition, we also gathered non-symptomatic sound events (e.g., door closing), which are labeled as "other" categories.

To test the performance of SymRecorder, a prototype is developed and installed on Honor-10 and Xiaomi 12Pro smartphones. Four volunteers who participate in data collection are joined by an additional 6 volunteers for evaluation purposes. The evaluation scenarios included home, office, canteen, and shopping mall. Over the course of nearly three months of evaluation, we collect 1331 cough, 797 sneeze, 916 t-c, and 1054 sniffle samples. Table 1 presents detailed information about the utilized dataset. We compare the performance of SymRecorder with the following methods, which also focus on detecting respiratory symptoms through acoustic sensing:

SymDetector [12]: This work classifies cough, sneeze, sniffle, and t-c symptoms using the SVM classifier using time-domain features and frequency-domain features such as symptom length, the center of mass, bandwidth, etc.

SymListener [16]: This work uses MFCC and GFCC features to classify cough, sniffle, and sneeze using Long Short Term Memory (LSTM) networks.

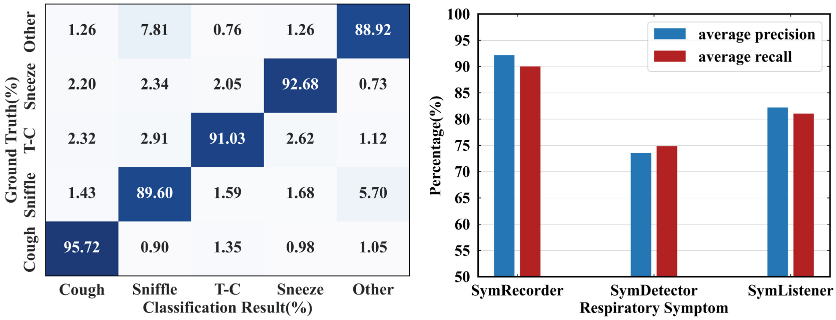
3.2 System Performance

Overall Performance. We first compare the overall performance of SymRecorder with the baseline methods realized in an offline manner. Figure 7a shows the confusion matrix of SymRecorder, indicating that 93.18% of respiratory symptoms are correctly classified. Sniffle has a probability of being classified as "other", but is less likely to be classified as cough. Cough has a probability of being classified as t-c, while sneeze has a probability of being classified as sniffle. Figure 7b illustrates the overall performance of SymRecorder compares to the two baseline methods. It can be observed that SymRecorder achieved the highest average recall and precision, which are 92.17% and 90.04%, respectively. Due to SymDetector relying only on audio amplitude and RMS to detect sound-related events, it is less robust to the interference of noisy environments, such as canteens and malls. This may result in SymDetector missing sound events in noisy environments. Although SymListener can adapt to strong driving noise, it

Table 1. Setup of Datasets.

dataset	cough	sneeze	t-c	sniffle	days	source
train set	2013	1310	1764	2341	–	vocalsound
	2873	2008	2577	3135	120	1st–14th
testset	1331	797	916	1054	85	11th–20th

does not consider the impact of consecutive symptoms, treating them as individual occurrences. Furthermore, both SymDetector and SymListener do not differentiate the source of symptoms, and symptoms generated by other people also lead to overall performance degradation. For SymRecorder, the detection accuracy for cough and sneeze is relatively high. This can be attributed to the high energy density and long symptom duration associated with these two symptoms. In contrast, the detection accuracy for sniffle is relatively low due to its lower energy density and shorter symptom duration.

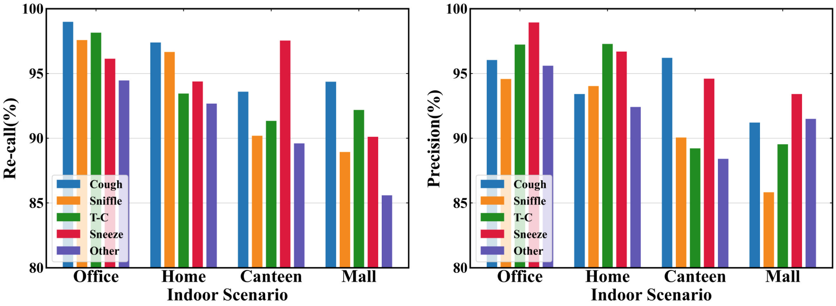


(a) confusion matrix

(b) overall performance

Fig. 7. The overall performance of SymRecorder.

Influence of Indoor Scenario. Figure 8a and Fig. 8b illustrate the recall and precision in different indoor scenarios. In this context, the term "mall" refers to a comprehensive commercial complex where the environmental noise tends to be more pronounced compared to other scenarios. It can be observed that SymRecorder performs the best in office environments, as offices are typically characterized by relatively quiet surroundings. Across various scenarios, the detection performance for cough and sneeze is consistently good. However, in canteen and mall scenarios, the recall and precision for sniffle are relatively low. This is because these scenarios often feature short and high-frequency sound events such as tray handling noises and buzzing sounds, which can either mask sniffle sounds or be misclassified as sniffle. Additionally, the category of "other" sound events exhibits a lower recall rate but higher precision in the evaluation. This suggests that sound events tend to be classified as respiratory symptoms, while respiratory symptoms are difficult to classify as "other". This phenomenon may be attributed to the fact that certain sound events can generate acoustic characteristics similar to respiratory symptoms.



(a) recall

(b) precision

Fig. 8. The performance of different scenarios.

3.3 Discussion and Future Work

Although we introduce a subdivision algorithm to handle potential instances of continuous cough events, the subdivision algorithm cannot handle all continuous cough events. The algorithm can fail when the second burst stage of a cough symptom resembles the first stage. Furthermore, if multiple individuals cough simultaneously, causing overlapping cough sounds, the subdivision algorithm may also produce errors. We will subsequently improve the subdivision algorithm and consider acquiring dual-channel signals from headphones to distinguish between different users.

4 Related Work

The audio-based approach has an excellent track record in detecting respiratory health. PulmoTrack-CC [14] achieved 94% overall specificity and 96% overall sensitivity in detecting cough events. VitaloJAKTM [2] proposes to capture signal regions with high energy and high spectral mass to automatically count coughs from recordings. However, all of the above work requires the user to wear a recording device or acoustic sensor, which is extremely inconvenient to use.

In many previous works, smartphones started to be used to collect respiratory health information. iSleep [8] is a smartphone-based sleep monitoring system that detects snoring sounds from the user, but it has high environmental requirements. symDetector [12] is a smartphone-based application that detects sneeze, cough, sniffle, and t-c sounds in a home or office environment, and has high ambient noise requirements. SymListener [16] is also a smartphone-based application that detects sneeze, cough, and sniffle sounds in the driving environment, with a high level of environmental robustness, but without considering the effects of continuous symptoms.

5 Conclusion

We propose SymRecorder, an application based on the microphone of earphones, which can inconspicuously detect user-related respiratory symptoms in various indoor environments, including cough, sneeze, t-c, and sniffle. A method called RA-ABSE is designed to detect the endpoints of sound events, and Berouti power spectral subtraction is employed to remove potential environmental noise. We devise an algorithm to subdivide possible continuous symptoms, utilizing MFCC, GFCC, and spectrogram as features, and employ the ResNet with the stacked attention mechanism and MLP for classification. Extensive experiments are conducted to evaluate the performance of SymRecorder in different indoor environments, and the results demonstrate that SymRecorder can detect respiratory symptoms with high accuracy.

References

1. Akhil, S., et al.: A novel approach for detection of the symptomatic patterns in the acoustic biological signal using truncation multiplier. In: ICICICT 2019, pp. 49–53 (2019). https://doi.org/10.1109/ICICICT46008.2019.8993389
2. Barton, A., Gaydecki, P., Holt, K., Smith, J.A.: Data reduction for cough studies using distribution of audio frequency content. Cough **8**, 12 (2012). https://doi.org/10.1186/1745-9974-8-12
3. Wu, B.-F., Wang, K.-C.: Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments. IEEE Trans. Speech Audio Process. **13**, 762–775 (2005). https://doi.org/10.1109/TSA.2005.851909
4. Chauhan, J., Hu, Y., Seneviratne, S., Misra, A., Seneviratne, A., Lee, Y.: BreathPrint: breathing acoustics-based user authentication. In: MobiSys 2017, pp. 278–291 (2017). https://doi.org/10.1145/3081333.3081355
5. Chung, K.F., et al.: Cough hypersensitivity and chronic cough. Nat. Rev. Dis. Primers. **8**, 45 (2022). https://doi.org/10.1038/s41572-022-00370-w
6. French, C.T., Irwin, R.S., Fletcher, K.E., Adams, T.M.: Evaluation of a cough-specific quality-of-life questionnaire. Chest **121**, 1123–1131 (2002). https://doi.org/10.1378/chest.121.4.1123
7. Gong, Y., Yu, J., Glass, J.: Vocalsound: A dataset for improving human vocal sounds recognition. In: ICASSP 2022, pp. 151–155 (2022). https://doi.org/10.1109/ICASSP43922.2022.9746828
8. Hao, T., Xing, G., Zhou, G.: iSleep: unobtrusive sleep quality monitoring using smartphones. In: Sensys 2013, pp. 1–14 (2013). https://doi.org/10.1145/2517351.2517359

9. Korpáš, J., Sadloňová, J., Vrabec, M.: Analysis of the cough sound: an overview. Pulm. Pharmacol. **9**, 261–268 (1996). https://doi.org/10.1006/pulp.1996.0034
10. Lu, H., Pan, W., Lane, N.D., Choudhury, T., Campbell, A.T.: SoundSense: scalable sound sensing for people-centric applications on mobile phones. In: MobiSys 2009, Kraków, Poland, pp. 165–178 (2009). https://doi.org/10.1145/1555816.1555834
11. Qian, K., et al.: Acousticcardiogram: monitoring heartbeats using acoustic signals on smart devices. In: INFOCOM 2018, pp. 1574–1582 (2018). https://doi.org/10.1109/INFOCOM.2018.8485978
12. Sun, X., Lu, Z., Hu, W., Cao, G.: SymDetector: detecting sound-related respiratory symptoms using smartphones. In: UbiComp 2015, pp. 97–108 (2015). https://doi.org/10.1145/2750858.2805826
13. Vhaduri, S., Kessel, T.V., Ko, B., Wood, D., Wang, S., Brunschwiler, T.: Nocturnal cough and snore detection in noisy environments using smartphone-microphones. In: ICHI 2019, pp. 1–7 (2019). https://doi.org/10.1109/ICHI.2019.8904563
14. Vizel, E., et al.: Validation of an ambulatory cough detection and counting application using voluntary cough under different conditions. Cough **6**, 3 (2010). https://doi.org/10.1186/1745-9974-6-3
15. Wang, C., Peng, J., Song, L., Zhang, X.: Automatic snoring sounds detection from sleep sounds via multi-features analysis. AUST. Phys. Eng. Sci. **40**, 127–135 (2017). https://doi.org/10.1007/s13246-016-0507-1
16. Wu, Y., Li, F., Xie, Y., Wang, Y., Yang, Z.: SymListener: detecting respiratory symptoms via acoustic sensing in driving environments. ACM Trans. Sens. Netw. **19**, 1–21 (2023). https://doi.org/10.1145/3517014
17. Xie, Y., Li, F., Wu, Y., Wang, Y.: HearFit: fitness monitoring on smart speakers via active acoustic sensing. In: INFOCOM 2021, pp. 1–10 (2021). https://doi.org/10.1109/INFOCOM42981.2021.9488811
18. Xie, Y., Li, F., Wu, Y., Yang, S., Wang, Y.: D^3-guard: acoustic-based drowsy driving detection using smartphones. In: INFOCOM 2019, pp. 1225–1233 (2019). https://doi.org/10.1109/INFOCOM.2019.8737470
19. You, M., et al.: Novel feature extraction method for cough detection using NMF. IET Signal Process. **11**, 515–520 (2017). https://doi.org/10.1049/iet-spr.2016.0341