

On the Use of Bronchial Breath Sounds for Person Identification*

VAN-THUAN TRAN, YUNG-CHIH LIN AND WEI-HO TSAI

*Department of Electronic Engineering
National Taipei University of Technology
Taipei, 106 Taiwan*

E-mail: thuan.tranvan586@gmail.com; t103368046@ntut.edu.tw; whtsai@ntut.edu.tw

Bronchial breath sound is the sound of turmoil flow produced by the inspiratory air through glottis, trachea or major bronchi. It cannot be only used to diagnose the respiratory tract and lung-related diseases but also used to distinguish one person from the other and thereby identifying patients since it contains personal physiological characteristics. This study captures the bronchial breath sound by using a stethoscope attached on a subject's neck. For each person, the Mel Frequency Cepstral Coefficients (MFCCs) are computed for his/her bronchial breath sounds, and then represented by a stochastic model. Given an unknown breath sound recording, the proposed person identification system determines who among a set of candidate people produced the breath sound by matching the MFCCs of the sound to each of the stochastic models. Furthermore, we apply the *i*-vector approach in the system to boost the identification accuracy. To evaluate the generality of our experimental results, we additionally utilize other general identification schemes including support vector machine, random forest, and naive Bayes. Our experiments conducted on a dataset composed of 8 persons show that the accuracy of identifying people from their breath sounds can attain 92%.

Keywords: bronchial breath sound, Gaussian mixture model (GMM), *i*-vector, MFCCs, person identification

1. INTRODUCTION

Person identification (PID) has long been an important need in many applications involving different treatments for different individuals. Biometric-based PID, which identifies an individual based on his/her distinguishing physiological and/or behavioral characteristics, has advantages on better convenience and reliability than the traditional token or knowledge-based identification, such as ID card or password. Thus, a great deal of research has been done on biometric-based PID techniques, such as facial imaging, iris or retina biometrics, signature, voice, finger/palm-print imaging, and finger/hand geometry. However, no single technique is absolutely superior to others, since each technique has its own occasion of use. For example, voice-based PID shows its superiority over fingerprinting and eye-based methods when hand-busy or eye-busy situations like driving are involved, but its performance may suffer from noisy environments. Thus, several biometric-based PID techniques sometimes may work together and even work with token or knowledge-based PID for double confirmation. In contrast to the existing PID techniques, this work proposes a novel biometric based on bronchial breath sounds for PID. To the

Received November 20, 2018; revised July 12, 2019; accepted July 24, 2019.

Communicated by Wen-Liang Hwang.

* This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant No. MOST 106-2221-E-027-125-MY2.

best of our knowledge, no prior research has been done on PID using *bronchial* breath sounds.

Breath sound is a result of ventilation caused by the expansion and contraction of the lungs. More specifically, the air flow that causes vibration in the trachea is defined as the breath sound, which can be heard with the aid of a stethoscope. In the past, breath sounds were commonly used by doctors for diagnoses. Recently, analyses of breath sounds have created new applications. For example, R. Darooei proposed a method to detect asthma by analyzing the length of breathing sounds, respiratory rate, and volume [1]. E. Kaniusas recorded sleep breathing sounds through a stethoscope to analyze whether breathing is aborted [2]. R. L. Moedomo *et al.* analyzed breathing rate per minute and exhaled carbon dioxide concentration to determine the degree of psychological stress [3]. In [4], Y. Ren introduced a system that has the capability to monitor an individual's breathing rate as well as sleep events using off-the-shelf smartphones. The system employs the readily available earphone for smartphones to capture the breathing sound and measure the breathing rate, and support vector machine is used as a classifier to detect and identify each sleep event including snore, cough, turn over, and get up. In [5], the authors applied semi-supervised deep learning algorithm for automatically classifying the lung sound of patients to diagnose lung-related diseases, which can be applied in remote patient monitoring application and point-of-care diagnostics. S. A. Taplidou *et al.* proposed a wheeze detection system, namely time-frequency wheeze detector (TF-WD) [6], which analyses the time-frequency characteristics of the breath sound signal to identify wheezing-episodes from breath sound recordings.

Breath sounds can be divided into two categories, namely, vesicular breath sounds and bronchial or tracheal breath sounds. Several mechanisms of vesicular breath sound generation are discussed in [7] and it is shown that turbulence in the central airways is the source of most vesicular sounds. Vesicular breath sounds are present at sites that are at a distance from large airways, and they are commonly heard over the chest. The respiratory sounds heard in the chest wall undergo the attenuation since the small airways, lung parenchyma and chest wall act as a low-pass filter which does not allow high-frequency components of the sound to pass through [8, 9]. Therefore, the vesicular breath sounds heard over the chest wall mainly consists of low-frequency components. In addition, the inhalation phase of vesicular breath sound is approximately two times longer than the exhalation phase during tidal breathing, and there is no pause between inhalation and exhalation [9]. Bronchial breath sound refers to the air that produces a wheezing sound between the glottis and the trachea. When breathing in, the glottis widens to allow the air to pass faster, whereas the glottis narrows as air flow slows down. Bronchial breathing sound can be heard nearby the tracheal position, throat, sternum, suprasternal fossa, and posteriorly between the 7th cervical vertebrae (C7) and the 3rd thoracic vertebrae (T3). Bronchial breath sounds contain more high-frequency components than vesicular breath sounds due to alteration of the low pass filtering function of the alveoli, as occurs in consolidation [8, 9]. In contrast to vesicular breath sound, bronchial breath sound is loud, hollow, and high pitched.

In this study, we choose bronchial breath sounds as the basis for PID given the difficulty of measuring vesicular breath sounds. In addition, the lower frequency distribution of vesicular breath sounds, which overlaps partially with heart sounds also result in difficulty in the follow-up treatment. By contrast, louder bronchial breath sounds with higher frequency distribution is advantageous to perform analysis.

Breath sound based PID is particularly suitable for confirming patients' identities in

hospitals. Without extra equipment, breath sound based PID could help doctors identify patients by simply using stethoscopes to prevent misdiagnosis. In some situations, that people are unconscious, such as sleeping, fainted, or anesthetized, breath sound based PID may be more useful than other biometric-based PID, such as voices or iris.

The remainder of this paper is organized as follows. In Section 2, we describe a fundamental of breath sound based person identification. Section 3 presents methods to implement a breath sound based PID system using Gaussian mixture model (GMM), *i*-vector in combination with linear discriminant analysis (LDA), support vector machine (SVM), naive Bayes (NB), and random forest (RF). Section 4 discusses different experiment scenarios and results. Then, we present our conclusions and indicate the direction of our future work in Section 5.

2. FUNDAMENTAL OF PERSON IDENTIFICATION BASED ON BRONCHIAL BREATH SOUNDS

2.1 Measurement of Bronchial Breath Sounds

Given an unknown breath sound recording, our aim is to develop a system that automatically determines who among a set of people produced the breath sounds. In this work, bronchial breath sound is recorded using a traditional stethoscope, CK-T601P, produced by Spirit Medical Co., Ltd. together with an audio-technica AT9931PC microphone. One earpiece of the stethoscope is connected to the microphone as shown in Fig. 1, so that bronchial breath sounds sensed from the stethoscope can be recorded into a computer with a microphone. The chest-piece of the stethoscope is attached to the subject's neck as in Fig. 2.



Fig. 1. Bronchial breath sound measurement tools: stethoscope and microphone.



Fig. 2. Position of bronchial breath sound measurement.

2.2 Bronchial Breath Sound Characteristics

In our experiments, the position near the trachea on the left side of the neck is selected for measurement of bronchial breath sounds, as shown in Fig. 2. Since the recorded breath sounds are produced close to the tracheal position, the sounds are louder and less overlapped with pulse tones. It is also indicated in [10] that we can obtain the largest power of breath sound signal when the sensor is attached on the neck. From Fig. 3 we can observe that there

are several pulse tones mixed into the waveform of recorded breath sound, these pulses are brought about by the cardiac vasomotor activity where blood flow is transmitted through the artery and causes the arterial wall to pulsate. In the waveform, the lower amplitude part from 0 to 2 seconds is the breath sound signal during the inspiratory phase, and the higher amplitude part from 2 to 4 seconds is the signal of the expiratory phase. Due to the difference in the physiological structure of each person, there is no absolute relationship between the amplitude and the phase of breath signals during inspiratory and expiratory phases. The subject producing the signal in Fig. 3 has significantly higher exhaled sound amplitude than inhaled sound; however, we observed the opposite trend in breath sound waveforms of several other subjects.

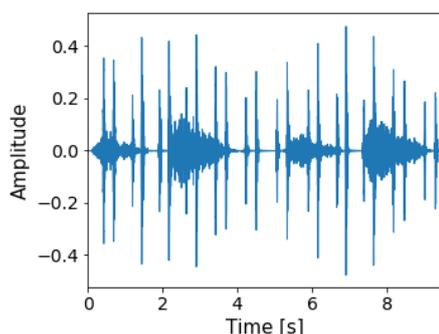


Fig. 3. The waveform of a bronchial breath sound recording.

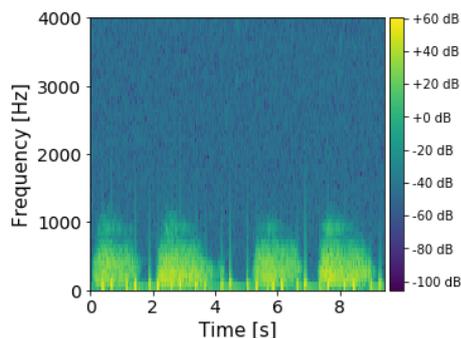


Fig. 4. Spectrogram of the bronchial breath sound in Fig. 3.

Fig. 4 shows the corresponding spectrogram of the breath sound signal in Fig. 3, in which the brighter color represents the higher energy. At the low-frequency range of 0-150 Hz, we can see the spectrogram of pulse sounds shown as the periodic pulses, whereas the remaining lighter color part is spectrogram of breath sound signal, indicating that most of pure breath sound energy is distributed between around 150 Hz and 750 Hz. In fact, some subjects can produce breath sounds that have the main frequency components above 850 Hz. Since the characteristics of bronchial breath sound are different from person to person, it is possible to employ breath sounds as the basis for identity recognition.

2.3 Different Breathing Ways

This study discusses several ways of breathing, mainly including guided breathing and natural breathing. In guided breathing, we regulate a subject's breathing on the basis of 2-second inhalation and 3-second exhalation. On the other hand, in natural breathing type subject breathes in a usual manner. In our experiments, we further consider three types of breathing:

- Type 1: Inhalation with nose and exhalation with nose (nasal breathing).
- Type 2: Inhalation with mouth and exhalation with mouth (mouth breathing).
- Type 3: Inhalation with nose and exhalation with mouth (nasal inhale oral exhale).

We requested participants to provide breath sound recordings taken by mentioned breathing types to analyze and explore the characteristics of corresponding signals. Firstly, we analyzed the signal of nasal breathing type which is referred to as Type 1 breathing. Figs. 5 and 6 respectively present the waveform and spectrogram of a nasal breathing sound. The prominent pulses in the time domain waveform are pulse tones from the carotid artery. The first 0 to 1.5 seconds is the nasal inhalation waveform while the latter 2.5 to 4.0 seconds is the signal of the exhalation phase. In the frequency domain, we can observe that the energy of the nasal breathing sound is concentrated within 1250 Hz, and there is no energy distribution from 750 to 800 Hz.

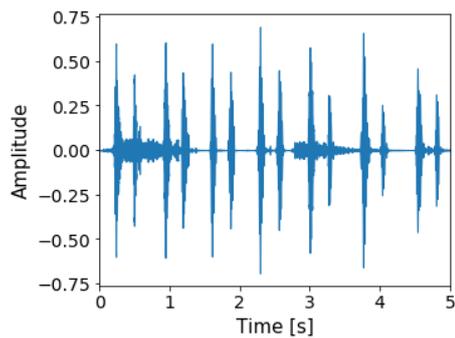


Fig. 5. The waveform of a nasal breathing sound recording.

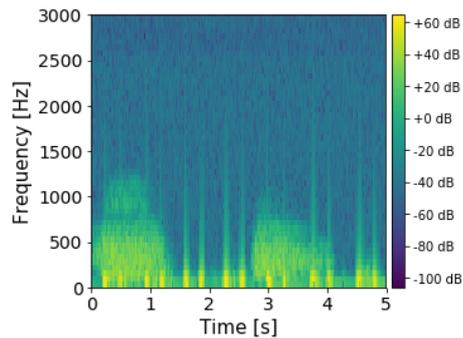


Fig. 6. Spectrogram of the nasal breathing sound in Fig. 5.

The second type of respiration is mouth breathing referred to as Type 2 breathing. The representations of mouth breathing sound in the time domain and frequency domain are shown in Figs. 7 and 8. In the time domain, it can be seen that the mouth breathing sound is clearer with larger amplitude compared to nasal breathing sound. In the frequency domain, most energy of both inhalation and exhalation phases is concentrated below 1000 Hz, in which from 100 to 750 Hz is the area where the energy distribution is mostly concentrated while from 750 to 800 Hz the energy is relatively small.

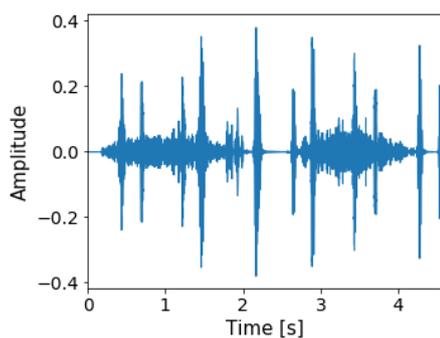


Fig. 7. The waveform of a mouth-breathing sound.

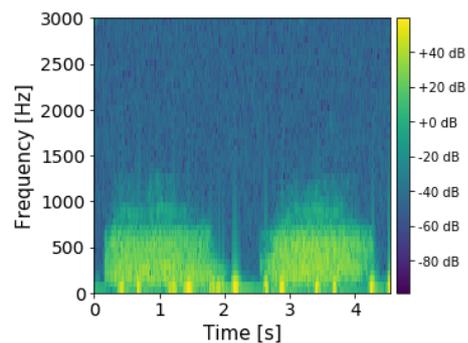


Fig. 8. Spectrogram of the mouth breathing sound in Fig. 7.

The third type of respiration is “nasal inhale and oral exhale” referred to as Type 3 breathing. Time-domain waveform and spectrogram of type 3 breath sound are shown in Figs. 9 and 10, respectively. It is obvious that the time of nasal inhalation is longer than that of oral exhalation. This is due to the relationship between the size of the nasal/oral aperture and speed of inhalation/exhalation: the larger the size of the oral aperture, the faster the exhalation speed. In the frequency domain, we can see that there is an obvious difference in energy distribution between inhalation/exhalation phases, resulting from a fact that the amount of exhaled air in the mouth is large so the energy is larger than that of nasal inhalation.

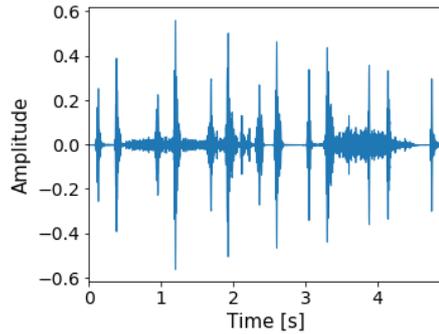


Fig. 9. The waveform of a type-3 (nasal inhale oral exhale) breath sound.

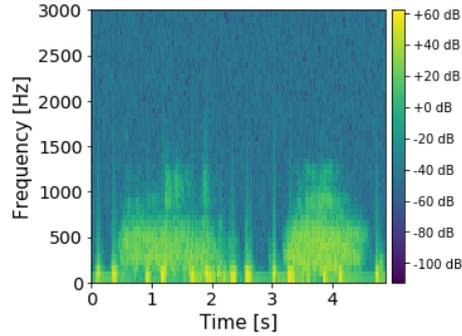


Fig. 10. Spectrogram of the type-3 (nasal inhale oral exhale) breath sound in Fig. 9.

2.4 Experimental Data Collection

Twenty subjects are invited to participate in our collection of experimental breath sound data. There was an equal number of female and male participants. The dataset is divided into two subsets. The first one, labeled as Dataset A, includes eight subjects, and the second one, labeled as Dataset B, includes the remaining twelve subjects. The data are collected on two discontinuous days for each subject, in which the data collected on the first day is labeled as Dataset A-I and Dataset B-I, and the data collected on the second day is labeled as Dataset A-II and Dataset B-II, respectively. On the second day of collection, we asked participants to provide data in different activities involving walking, running for a short while, and after climbing the stairs. For each type of breathing, every person provides 32 breath sound recordings. Each recording contains at least two breath cycles, up to 10 seconds in length.

3. METHODOLOGY

3.1 Identification Schemes

This study applies five classifiers to identify bronchial breath sounds, namely, Gaussian mixture model (GMM), *i*-vector approach, support vector machine (SVM), random forest (RF), and naive Bayes (NB).

3.1.1 GMM approach

Gaussian mixture models approach has become the widely used approach for modeling in text-independent speaker recognition applications. A GMM is used in speaker recognition applications as a generic probabilistic model for multivariate densities capable of representing arbitrary densities, which makes it well suited for unconstrained text-independent applications [11]. Because the breath sound may be also considered as a type of human voice, this study proposes using the Gaussian mixture model to evaluate the feasibility and effectiveness of breath sound identification. The process of identification can be roughly divided into two parts including training and testing as shown in Fig. 11. The training phase consists of feature parameters acquisition and identity model establishment while the testing phase is mainly separated into the feature extraction process of the unknown audio and model comparison followed by scoring step to make the final decision.

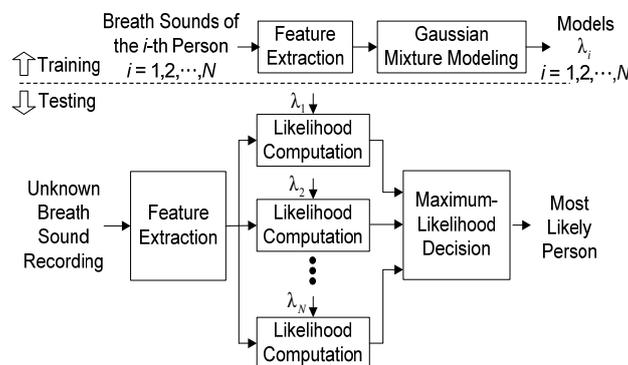


Fig. 11. Breath sound based PID using GMM approach.

Like human voice signal, breath sound signal is also time-varying and changes rapidly over time, which is not conducive to analysis. In contrast, in the frequency domain, the signal changes slowly with time. Therefore, the frequency domain analysis is more favorable than the time domain analysis for further processing steps. The parameters commonly used in speech signal analysis include Mel-frequency Cepstral Coefficients (MFCCs) and Linear Predictive Coding (LPC), in which the Mel-frequency system considers the nature of the human auditory system. In this study, we choose MFCC feature extraction method to derive the features of breath sound signals. Given a signal, the system partitions the signal into frames, invokes a window function to increase the continuity of voice signals in a frame, utilizes the Fast Fourier Transform (FFT) to convert the digital signals into spectrum data, and employs the triangular band-pass filter designed to simulate the spectral data of the human hearing. Finally, to obtain MFCCs, a Discrete Cosine Transform (DCT) is applied to the filter banks retaining a number of the resulting coefficients while the rest are discarded.

3.1.2 *i*-vector approach

This study also attempts to use *i*-vector in combination with linear discriminant analysis (LDA) to implement a breath sound based PID system. *i*-vector is a front-end factor analysis

method proposed by Dehak *et al.* [12] and it has been effectively applied to the speaker verification application [13, 14]. *i*-vector is inspired by the earlier use of joint factor analysis (JFA); however, unlike JFA which decomposes the speaker-dependent GMM super-vector into the separate speaker- and channel-dependent parts, *i*-vector uses only a single space called total variability space to represent GMM super-vector. The channel space of the JFA is still informative and its information can be used for classifying classes; therefore, the use of a total variability space can contribute to enhancing the identification capability of our system. Since the *i*-vector extraction algorithm does not separate the speaker variability and the channel variability, channel compensation methods such as LDA, within-class covariance normalization (WCCN) and/or nuisance attribute projection (NAP) should be utilized.

Fig. 12 is the diagram of a breath sound identification system based on the *i*-vector approach. The system can be divided into three parts including pre-processing and *i*-vector extraction, channel compensation, and identification score calculation. Firstly, pre-processing is carried out to retrieve breath sound feature parameters, filter noise, and normalize parameters. After that, we train the universal background model (UBM) and the total variability space R before extracting *i*-vectors. After obtaining *i*-vectors, LDA channel compensation technique is utilized to not only reduce the vector dimension but also to better distinguish between classes. Lastly, we employ the cosine distance scoring method to compute the decision score and make the identification decision.

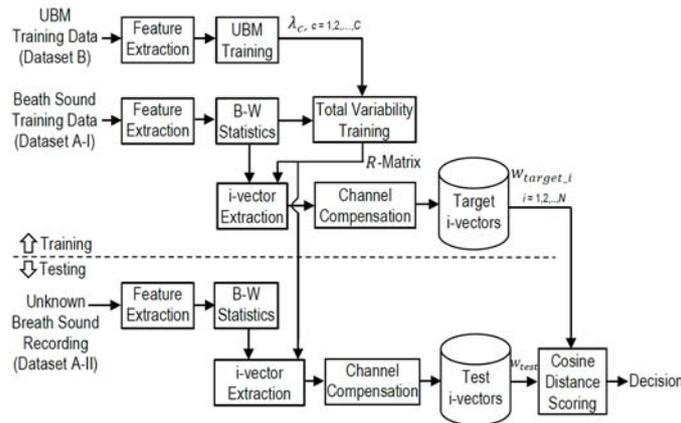


Fig. 12. Breath sound based PID using *i*-vector approach.
B-W: Baum-Welch statistics.

(a) *i*-vector extraction [12]

Basically, in the *i*-vector framework, the person- and channel-dependent GMM super-vector M is defined by Eq. (1) with an assumption that M has a normal distribution with mean vector m and covariance matrix RR'

$$M = m + Rw \quad (1)$$

where m represents the person- and channel-independent super-vector which can be extracted from UBM super-vector trained on a large development dataset, R is a low-rank rectangular

matrix, w is a random vector which is normally distributed with parameters $N(0, I)$, and the components of the vector w are the total factors. w is referred to as i -vector.

In a standard i -vector system, the Baum-Welch statistics is employed to estimate the i -vectors for given sound files. The details of the total variability space training and i -vector extraction process is given in [12].

(b) Channel compensation using the LDA technique

LDA is a widely used approach that can project data from a high-dimensional space to a lower one. The main idea of LDA is to find a new space that improves the ability to distinguish classes, which means that in the new space the between-class variance can be maximized while within-class variance can be minimized. In this study, we apply the concept of LDA for breath sound based PID, that is, to reduce the variation of the same subject's breath sounds, and to increase the variation of the breath sounds from different subjects. The optimization problem of LDA can be defined by the ratio $J(v)$ which represents the amount of information ratio of the between-class variance and within-class variance.

$$J(v) = \frac{v^t S_b v}{v^t S_w v} \quad (2)$$

$$S_b = \sum_{s=1}^S (w_s - \bar{w})(w_s - \bar{w})^t \quad (3)$$

$$S_w = \sum_{s=1}^S \frac{1}{n_s} \sum_{i=1}^{n_s} (w_i^s - \bar{w}_s)(w_i^s - \bar{w}_s)^t \quad (4)$$

In Eq. (2), v is space direction; S_b and S_w are respectively the between-class and within-class variance matrices defined by Eqs. (3) and (4) where $\bar{w}_s = (1/n_s) \sum_{i=1}^{n_s} w_i^s$ is the mean of i -vectors for each subject, S is the number of subjects, and n_s is the number of breath sound files for each subject s .

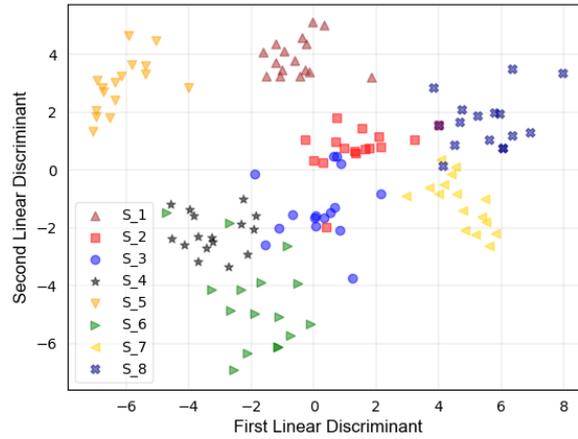


Fig. 13. Distribution of training data after applying LDA.

In LDA, we aim to maximize the value of $J(v)$ and find the projection matrix A composed by eigenvectors with the highest eigenvalues of the objective Eq. (5).

$$S_b v = \Lambda S_w v \quad (5)$$

where Λ is a diagonal matrix formed by eigenvalues. Projection matrix A attained from LDA is afterward used for linear conversion of i -vectors to the new space. Fig. 13 indicates the 2-dimensional distribution of training data after submitting i -vectors to the matrix A . In this figure, the breath sound samples from eight subjects are represented by eight different marks.

(c) Identification scoring

At the final step, the cosine distance scoring method is used to make the identification decision. In this method, we directly use the value of the cosine kernel between each target person's i -vector w_k , $1 \leq k \leq N$, and test i -vector w_{tst} as the basis for final decision making.

$$\text{score}(w_k, w_{tst}) = \frac{\langle w_k, w_{tst} \rangle}{\|w_k\| \|w_{tst}\|} \quad (6)$$

The most likely person K^* who produced the test breath sound is determined as the one satisfying

$$K^* = \arg \max_{1 \leq k \leq N} \text{score}(w_k, w_{tst}). \quad (7)$$

3.1.3 General classifiers: SVM, NB, and RF

In order to evaluate the generality of our experimental results, the results obtained from the stochastic model method (GMM) and i -vector approach are compared with several PID system based on SVM, naive Bayes, and random forest.

Support vector machine is one of the powerful classifiers applied efficiently to various classification problems, such as pattern recognition and speaker identification. The detailed description of SVM is given in [15, 16]. SVM is initially designed to deal with the binary classification problem. Therefore, in order to solve a problem with only two classes, we can directly apply a single SVM. However, to extend SVM for a multi-class classification problem we need to use a number of binary classifiers together with special techniques, namely "one-against-rest method" and "one-against-another method" or "pairwise method". In this study, we aim to identify breath sounds from eight participants, so it is considered as a multi-class classification problem in SVM. Moreover, in a study to compare the efficiency of "one-against-rest method" and "one-against-another method", C. W. Hsu and C. J. Lin found that the former is less effective than the later one [17]. As a consequence, we adopt the "one-against-another method" to tackle the multi-class classification problem in this study.

Naive Bayes is a supervised learning technique based on the Bayesian theory, and it is an efficient probabilistic classifier widely used in pattern recognition. In NB, we have a strong independent assumption that the feature items in one class are independent of other attribute values [18]. Since breath sound based PID is a multi-class classification problem which can be handled by Naive Bayes classifiers, NB is also used here for classifying the

breath sound signals. In addition, Naive Bayes classifier can work with a small training dataset to estimate the parameters of the classification system, this is another reason why NB can be potentially applied to our case of breath sound based PID with a moderate dataset. Besides, in this study, we apply another supervised classification algorithm called random forest which utilizes the ensemble approach to improve the classification performance. The fundamental idea of the random forest is to combine a number of decision trees into a single model; in other words, we group several weak learners to form a strong learner. In actual practice, RF is formed by a large number of decision trees ranging from tens to hundreds depending on the particular classification task.

3.2 Breath Sound Pre-processing

In practical use, breath sounds can be recorded in a noisy environment existing several sources of noise such as speech, machine on operating, door closing, and event noise generated by subjects. Under such condition, it is necessary to do pre-processing on our data to mitigate the effects of noise.

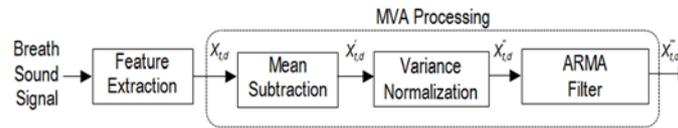


Fig. 14. MVA post-processing technique.

Fig. 14 is the diagram of the pre-processing method applied in this study. Firstly, we extract MFCC features, and then use the MVA technique [19] to generate the robust feature vectors. In feature extraction, we not only consider MFCC features but also consider their first-order and second-order derivatives which provide the information of dynamics of MFCCs over time. In fact, using the derivatives of MFCCs together with original MFCC features increases the performance in a number of audio analytics applications.

The MVA processing can be broken down into three parts including mean subtraction (MS), variance normalization (VN), and auto-regression moving average (ARMA) filtering as shown in Fig. 14. MS is used to normalize the first-order moment of the feature vectors. Suppose $X_{t,d}$ is the feature vector extracted from the noise-containing breath sound at time t , d represents the dimension of feature space, and u_d is the mean vector estimated from data. After the mean value is eliminated we get the subtracted feature $X'_{t,d}$ as in Eq. (9).

$$u_d = \frac{1}{T} \sum_{t=1}^T X_{t,d} \quad t = 1, 2, \dots, T \quad d = 1, 2, \dots, D, \quad (8)$$

$$X'_{t,d} = X_{t,d} - u_d. \quad (9)$$

The mean value of each dimension from the feature vector after MS step is equal to zero so that the feature parameters are less influenced by the environment at the time of extraction. MS can contribute to reducing the channel effects by removing time-invariant distortions introduced by the transmission channel and recording device [20]. The VN step further nor-

malizes the second-order moment of the feature vector. Assume σ_d is the standard deviation of the d -dimensional feature parameter, so the VN step is defined by Eq. (11).

$$\sigma_d = \sqrt{\frac{1}{T} \sum_{t=1}^T (X_{t,d} - u_d)^2} \quad (10)$$

$$X_{t,d}'' = \frac{X_{t,d}''}{\sigma_d} \quad (11)$$

After variance normalization, each dimension parameter not only has zero mean but also has a variance of one. In addition, to mean subtraction, VN contributes to further mitigate the influence of environment on the feature parameters.

The final step of MVA post-processing is processing by ARMA filtering. The ARMA filter is a low-pass filter used together with the MS and VN to achieve a good additive effect. The main function of ARMA filter is to reduce the feature parameter sequence and eliminate the problem that the parameters change too fast, so as to meet the requirement that the signal changes relatively slowly in a short time interval. The expression of ARMA is defined by Eq. (12), where L is the order of ARMA filter.

$$X_{t,d}''' = \frac{X_{(t-L),d}'' + \dots + X_{(t-1),d}'' + X_{(t+1),d}'' + \dots + X_{(t+L),d}''}{2L + 1} \quad (12)$$

4. EXPERIMENTS AND RESULTS

This section describes the experimental results of the breath sound based PID using Gaussian Mixture Model, i -vector approach and general identifiers. In order to find out the effects of several breathing ways and data preprocessing on breath sound identification, we divide our experiments into three scenarios for discussion. In our experiments, the identification accuracy is defined by:

$$\text{Accuracy(in \%)} = \frac{\#\text{Correctly Identified Recordings}}{\#\text{Testing Recordings}} \times 100\%.$$

4.1 First Experiment Scenario

In this experiment scenario, firstly we compare the influences of three breathing types on identification performance so that we can find the most appropriate one facilitating the identification purpose. Following breathing types are examined respectively: nasal breathing, mouth breathing and nasal inhale oral exhale. We collected data for this primary experiment by natural breathing method, in which participants breathe in their natural manner. Secondly, since bronchial breath sounds are recorded along with the pulse sounds, we perform an experiment of pulse sound identification to evaluate how the mixed pulses can contribute to identity recognition, which is the fundamental for other experiments afterward.

4.1.1 Breath sound identification across different breathing types

In this experiment, we performed PID on a database collected from three breathing types. There are 8 subjects participating in our collection of experimental data. In each type of breathing, each person provided 32 breath sound recordings consisting of 16 recordings (Dataset A-I) for training and the remaining 16 recordings for testing (Dataset A-II). We set the sampling frequency to 16 KHz, and the length of each file is about 5 to 6 seconds. 39-dimensional features (including 13-MFCCs as well as their first-order and second-order derivatives) are extracted from each recording to use as the input of our models. Since the frequency components of breath sounds are mostly smaller than 1500 Hz, we set the upper cut-off frequency to 2000 Hz. Table 1 shows the results of breath sound identification on three breathing types using GMM, random forest, SVM and naive Bayes classifiers.

Table 1. Results of breath sound identification across three breathing types.

Type of Breathing	Identification Approach and Accuracy			
	GMM	RF	SVM	NB
Nasal breathing	*72.65%	48.43%	64.06%	59.37%
Mouth breathing	62.50%	39.06%	56.25%	53.12%
Nasal inhale Oral exhale	70.31%	42.96%	61.71%	54.68%

* is the best value

From the detailed statistics in Table 1, it is clear that the identification accuracies of nasal breathing sounds are significantly higher than that of other breathing types. We achieved the highest identification accuracy of 72.65% with a dataset collected from nasal breathing type and classified by GMM approach. On the other hand, the figures for other breathing methods are much lower, especially for mouth breathing which obtains the lowest results fluctuating from 39.06% (RF approach) to 62.50% (GMM approach).

According to [21], nasal breathing is a normal, most common type of breathing and it has two major advantages over mouth breathing: filtration of particulate matter by the vibrissae hairs and better humidification of inspired gas. Therefore, in a natural manner, the prominent experimental result of nasal breathing sound is an advantage to our study as we can employ this common breathing type to collect data for the remaining experiments.

4.1.2 Pulse sound identification

The purpose of this experiment is to find out how recorded pulse sounds can influence the result of breath sound identification. As a consequence, the feasibility of using pulse sounds for identity recognition is taken into consideration. Basing on the result of this experiment, we can further try to examine whether the performance of breath sound based PID system can be improved with or without the presence of pulse sounds. Figs. 15 and 16 are respectively the waveform and spectrogram of pulse sound from a volunteer. The first pulse in the waveform is pulse tone when the heart contracts and the second pulse is signal during diastole. From Fig. 16, we can observe that the energy of pulse sound mostly distributes from 0-200 Hz.

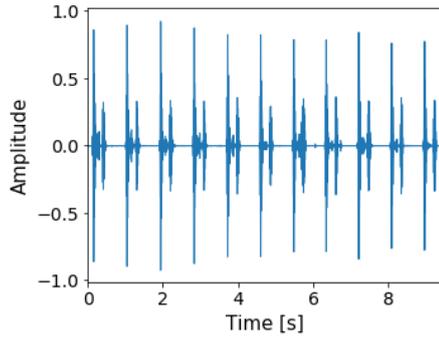


Fig. 15. The waveform of pulse sound from the first subject.

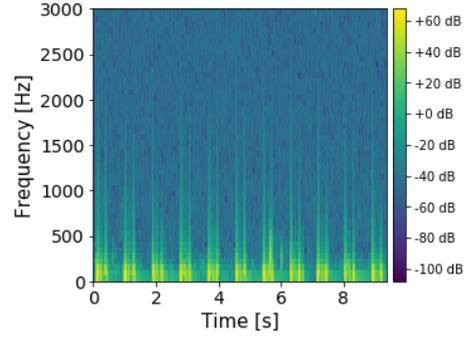


Fig. 16. Spectrogram of pulse sound from the first subject.

The number of participants in this experiment is 8 and 32 pulse sound recordings are provided by each person. The recordings in Dataset A-I are used for training our models while the remaining 128 recordings in Dataset A-II are used for testing, and each recording has a length of 9 to 10 seconds. In the feature extraction stage, 39-dimensional features consisting of MFCCs and their first-order and second-order derivatives are extracted from each recording.

We performed identification experiments using several approaches and obtained the highest accuracy of 39.06% with GMM method. RF, SVM, and NB also yielded similarly poor performances as shown in Table 2. Fig. 17 is the confusion matrix of pulse sound identification using GMM method. It is obvious that except subject 3 (S3) and subject 8 (S8), the identification accuracies are relatively low and most of the misidentification is concentrated on subject 1 (S1). In conclusion, it is not feasible to use pulse sound for person identification.

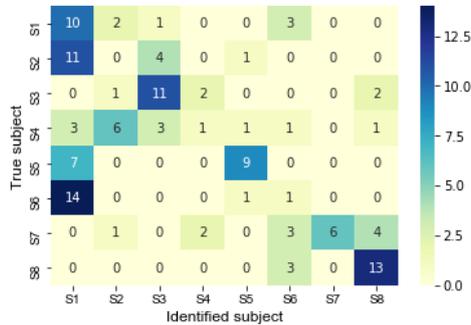


Fig. 17. Confusion matrix of pulse sound identification using GMM approach.

Table 2. Results of the pulse sound identification experiment.

Sound Type	Identification Approach and Accuracy			
	GMM	RF	SVM	NB
Pulse sounds	*39.06%	25.78%	9.375%	35.93%

* is the best value

4.2 Second Experiment Scenario

This experiment scenario focuses on exploring the influence of respiratory rhythm and energy factor on breath sound identification. Firstly, we analyze the energy characteristics of breath sounds from two datasets recorded by two breathing manners including natural breathing and guided breathing. After that, we perform independent experiments on those datasets. In the part of natural breathing, there is no restriction in breathing patterns, the frequency of respiration varied according to the individual's natural habits. Each recording has a length of about 9-10 seconds that is equivalent to 2-4 breath cycles. On the other hand, the guided breathing sounds are recorded by asking subjects to breathe deeply in accordance with the rhythm of approximately 2-second inhalation and 3-second exhalation; each recording is also about 9 to 10 seconds in length, approximately 2 breath cycles.

According to the results obtained from the first experiment scenario, the identification effectiveness of nasal breathing sounds is most prominent. Therefore, in this experiment scenario, we only asked participants to provide nasal breathing sounds but the length of each recording is longer. Furthermore, apart from using GMM which yielded the best results in the first experiment scenario, we also employed the *i*-vector approach and compared their identification performances.

4.2.1 Energy analysis of natural-breathing sounds and guided-breathing sounds

We analyzed the changes in the energy of bronchial breath sounds from each subject in order to find out the characteristics which contribute to improving the identification performance. Fig. 18 is the energy analysis of natural breathing sounds, where the horizontal axis represents the subject's index while the vertical axis represents the average energy of the subjects' breath sounds. The values of average energy (AE) and energy standard deviation (ESD) of natural breathing sounds from each subject are provided in a table at the bottom side. Likewise, Fig. 19 shows the energy analysis of the guided breathing sounds.

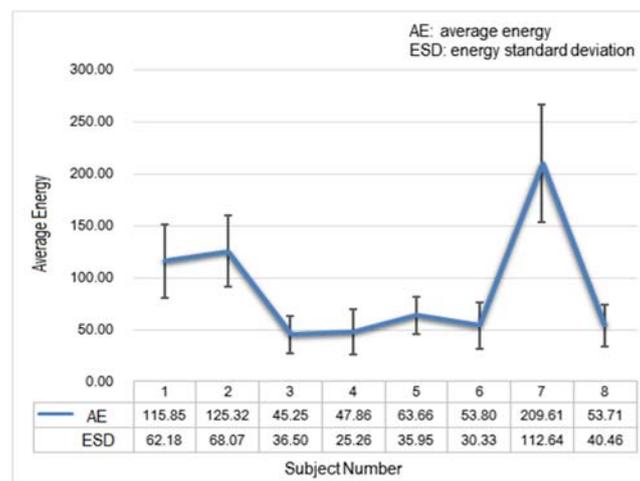


Fig. 18. Energy analysis of natural breathing sounds.

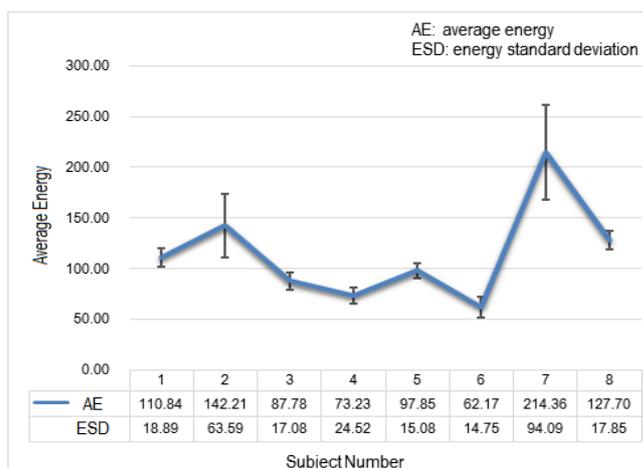


Fig. 19. Energy analysis of guided breathing sounds.

In comparison, the average energy of guided breathing sounds is larger than that of the natural breathing sounds, and the energy standard deviation of the former is much smaller than the figures of the latter. This indicates that the variation of the energy between guided breathing sounds is smaller, in other words; the change in energy between the natural breathing sounds is larger. This is mainly because the rhythm of guided breathing is regulated, in which subjects are required to take deep breath with the rule of 2-second inhalation and 3-second exhalation. Thus, guided breathing sound has larger average energy and there is an obvious downward change in the values of energy standard deviation. In addition, there is also a certain gap between average energy from subject to subject, so during the preprocessing stage, we also include the energy parameters (Log Energy) in the scope of feature extraction.

4.2.2 Identification experiments

The breathing type adopted in this experiment scenario is nasal breathing only. We trained the models basing on MFCC-based features together with energy parameters extracted from breath sound recordings. Table 3 summarizes the results of the second experiment scenario.

Table 3. Results of the second scenario experiment.

Breathing Method	Identification Approach and Accuracy	
	GMM	<i>i</i> -vector
Natural breathing	75.00%	79.68%
Guided breathing	87.50%	*89.84%

* is the best value

Firstly, we conducted identification experiments with GMM approach. In the natural breathing part, we tried to identify 128 breath sound recordings from 8 subjects, 96 of which are correctly identified, and the overall identification accuracy is 75%. It is indicated by the

statistics in Fig. 20 that the lowest identification rate is from the 7th subject (S7) whose natural breathing sounds has the highest energy standard deviation. In guided breathing sound identification, we achieved considerable improvement in system performance, in which 112 out of 128 recordings were identified correctly, resulting in the overall identification accuracy of 87.5%, 12.5% higher than natural breathing test. Fig. 21 indicates the detailed result of guided breathing sound identification using GMM, the identification accuracy of the 7th subject (S7) still stays the same at the lowest value of 50% as in natural breathing manner.

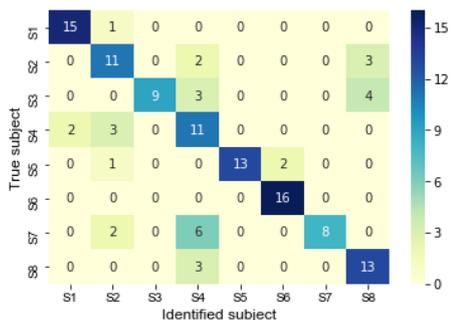


Fig. 20. Confusion matrix of natural breathing sound identification using GMM approach.

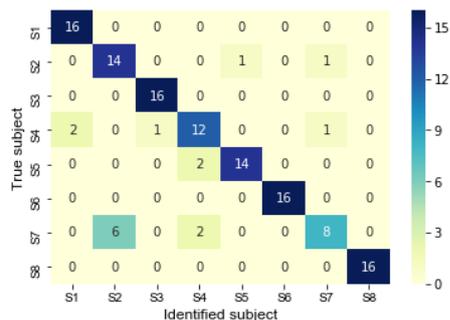


Fig. 21. Confusion matrix of guided breathing sound identification using GMM approach.

In the *i*-vector experiment, we also carried out identification with natural and guided breathing datasets separately. Before the establishment of the *i*-vector model, a universal background model is trained to serve as a reference for our model. We set the number of Gaussian components to 64 to train UBM model on 1024 breath sound recordings from Dataset B provided by the 12 subjects. There are 128 natural breathing sound recordings in the testing set (Dataset A-II), 102 of which are correctly identified, and the overall identification accuracy is 79.68%. Among all subjects, the identification rate of the second subject (S2) is lowest, at 43.75% as shown in Fig. 22. On the other hand, we achieved higher accuracy of 89.84% from the experiment on guided breathing sound dataset. Generally, the identification accuracies for almost subjects are significantly improved, as shown in Fig. 23.

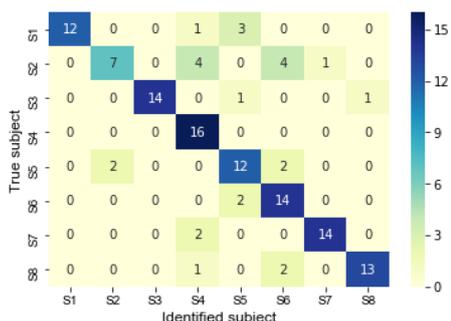


Fig. 22. Confusion matrix of natural breathing sound identification using *i*-vector approach.

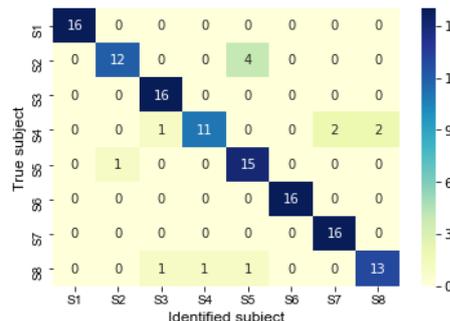


Fig. 23. Confusion matrix of guided breathing sound identification using *i*-vector approach.

According to results from this experiment scenario, we can conclude that the identification accuracy of guided breathing sounds is higher than that of natural breathing sounds, as indicated by statistics in Table 3. Our results are consistent with the initial assumption that the smaller standard deviation of energy the better capability of identification. In fact, during the collection of natural breathing sounds some subjects could provide small breath sounds resulting in an inaccurate sound collection and affects the identification performance. In contrast, in guided breathing, by regulating subjects' breath rhythm and asking subjects to take deeper breath we can get a better experimental dataset leading to better experiment results.

Another conclusion is that *i*-vector approach yielded better performance than GMM counterpart. Using *i*-vector approach, we obtained identification accuracy of natural respiration at 79.68% which is higher than 75.0% of GMM method. Similarly, the identification accuracy of guided breathing sounds also increased from 87.5% (GMM approach) to 89.84% (*i*-vector approach).

4.3 Third Experiment Scenario

As per the result from the first experiment scenario, due to the poor performance of pulse sound identification, we can consider that mixed pulse tones in breath sound data may not contribute efficiently to the result of breath sound based PID and even negatively affect the system performance. Recognizing this, we further performed the identification of pure breath sounds to explore whether the system performance can be better by removing the pulse tones from original breath sound data.

Our measure to obtain a pure breath sound dataset is filtering out the pulse tones from the original recordings. In Section 2, we found that the frequency components of pulse sounds recorded at our neck is ranging from 0 to 150 Hz, so we filter out all frequency components smaller than 100 Hz, this cut-off frequency value of 100 Hz is chosen to avoid information loss resulted in from filtering process. Figs. 24 and 25 are waveform and spectrogram of a pure breath sound's file, we can see that only clean breath sound is presented in the waveform, and the frequency components below 100 Hz are clearly removed from the corresponding spectrogram.

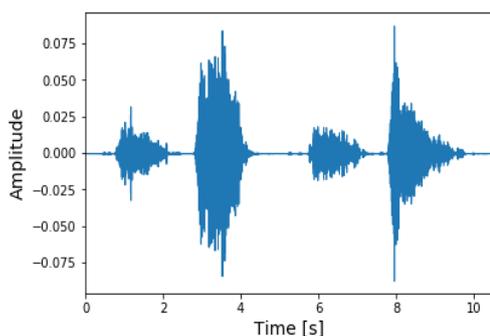


Fig. 24. The waveform of a pure breath sound.

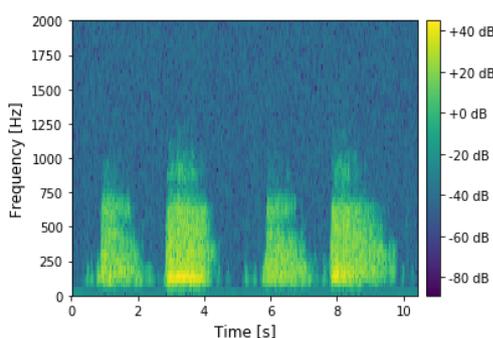


Fig. 25. Spectrogram of the pure breath sound in Fig. 27.

Due to the prominent results of identification on guided, nasal breathing sounds (nasal breathing, and breath rhythm is regulated) in previous experiment scenarios, in this experiment, we used the dataset of guided breathing sounds in second experiment scenario as the original dataset. Two experiments on pure breath sounds are performed using GMM and *i*-vector methods, respectively. The experiment configuration is mainly similar to the second experiment scenario except we additionally set the lower cut-off frequency to 150 Hz to filter out the pulse tones from our data. Table 4 shows the results of the third experiment scenario. In both identification approaches, the overall identification accuracies have increased when the pulse sounds are filtered out from original data.

Table 4. Results of pure breath sound identification experiment.

Identification Approach	Breath Sound Data	Accuracy
GMM	Original data	87.50%
	Filtered out pulse sounds	89.84%
<i>i</i> -vector	Original data	89.84%
	Filtered out pulse sounds	*92.97%

* is the best value

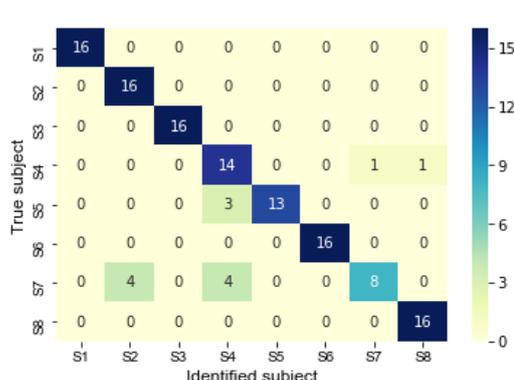


Fig. 26. Confusion matrix of pure breath sound identification using GMM approach.

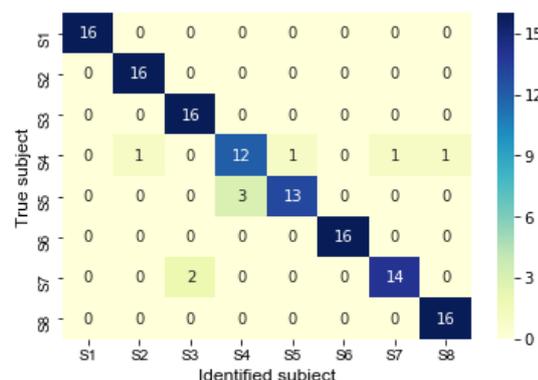


Fig. 27. Confusion matrix of pure breath sound identification using *i*-vector approach.

Fig. 26 is the confusion matrix of pure breath sound identification experiment using GMM approach. After filtering out the pulse sounds, the identification accuracy increases by 2.34% to 89.84% as there are 115 out of 128 sound files are correctly identified. Fig. 27 shows the result of pure breath sound identification obtained by *i*-vector approach. The accuracy increases from 89.84% (without filtering out pulse sounds) to 92.97%. The overall identification accuracy is improved significantly and among all subject the identification accuracies of subject number 2 (S2) and 8 (S8) increase considerably (compared to result in Fig. 23).

In conclusion, it is clear that identification performance can be improved by removing pulses tones from recorded breath sounds. We finally achieved the highest identification accuracy of 92.97% when pure breath sounds are identified by the *i*-vector system.

4.4 The Influence of Background Noise on System Performance

This experiment evaluates if the proposed identification system is capable of resisting to ambient noises. We collected two levels of noise signals. The first one represents the low level of noise signals, which were recorded in a normal condition of a business office. The second one represents high level of noise signals, which were collected when there are more activities in the office such as playing music on a mobile phone, having discussions, and printing documents. We used the same tools and procedures described in Section 2.1 to collect the background noise samples. Figs. 28 and 29 are spectrograms of a level-1 noise sample and a level-2 noise sample, respectively. For each noise level, we randomly chose a sample of background noise and added it to a sample of testing data in the guided breathing sound dataset, so that two new testing datasets corresponding to two levels of background noise were created. The experiment configuration in this analysis is the same as in the third experiment scenario, we test the system performance with the two datasets containing ambient noises.

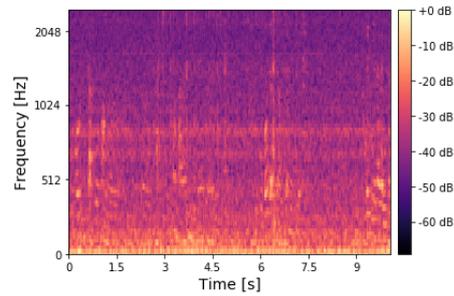
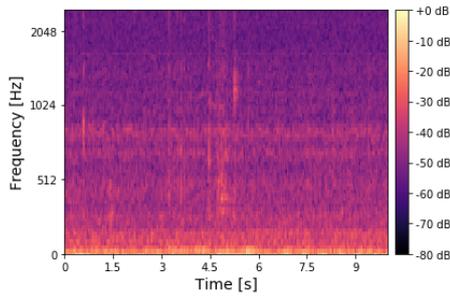


Fig. 28. Spectrogram of a level-1 noise sample. Fig. 29. Spectrogram of a level-2 noise sample.

Table 5. The influence of background noises on the system performance.

Identification Approach	Accuracy		
	Clean Data	Data with level-1 noise	Data with level-2 noise
GMM	89.84%	88.28%	86.72%
<i>i</i> -vector	*92.97%	92.97%	90.63%

* is the best value

Table 5 shows the results of this experiment. We can clearly see that the identification system quite resists to ambient noise. The system performance just decreases moderately, especially in the normal condition (level-1 noise) where the accuracies almost remain unchanged. In the noisier environment (level-2 noise), the identification accuracies are down slightly by 2.7%, from 92.97% and 89.84% to 90.63% and 86.72% for *i*-vector and GMM approaches, separately.

5. CONCLUSION AND FUTURE WORKS

This study has proposed a novel method for biometric-based person identification using bronchial breath sounds recorded from each person. Most of the previous breath sound

related studies are used to diagnose diseases or to determine the phase of breathing while this study used the measured breathing signals to achieve the identification purpose. We examined several factors that influence the identification performance, such as different types of breathing, breathing manners (natural breathing and guided breathing), the interference of pulse sounds and different identification methods. In this paper, three experiment scenarios were conducted to find out the most suitable condition for person identification based on bronchial breath sounds.

In the first experiment scenario, different breathing types were considered to assess their influences on the performances of identification systems. We divided the way of breathing into three types and then found the best one for breath sound identification application. Across all applied identification approaches, nasal breathing (inhalation with nose and exhalation with nose) yielded the highest identification accuracies, peaking at 72.65% with GMM identifier. In the second experiment scenario, we analyzed the energy distribution of natural breathing sounds and guided breathing sounds. We found that the energy of guided breath sounds is more stable with smaller standard deviation, resulting in better identification capability. In the natural breathing, the accuracy of GMM identifier is 75%, and that of the *i*-vector method is 79.68%. In guided breathing, we obtained the accuracies of 87.5% and 89.84% using GMM and *i*-vector approaches, respectively.

When we record the breath sound on the subject's neck by a microphone attached to an earpiece of a stethoscope, not only the pure breath sounds but also the pulse tones are recorded. Therefore, we further conduct the third experiment scenario to evaluate how the pulse sounds affect the identification results. In this experiment, firstly we carried out the PID test on pulse sounds dataset and evaluated the performance. We found that the identification accuracies are similarly low over all identification methods, so we decided to filter out the pulse sounds from the original data to obtain a clean experimental dataset. Eventually, we achieved the final identification accuracies of 89.84% and 92.97% on clean dataset identified by GMM and *i*-vector approaches, respectively. Moreover, among the applied identification methods, *i*-vector approach presented better performance.

The effect of background noise on the identification system has been examined also. We tested the system in two levels of background noise, it was shown that the proposed system has good robustness as its accuracy just slightly reduced in the noisy environments.

The results with the highest accuracy of 92.97% we obtained at this pilot investigation are encouraging and lay a good foundation for the future development of a person identification system based on bronchial breath sounds. However, future development is needed to meet the requirements of actual applications. First of all, in the future, we will extend the experimental dataset by taking more samples from a larger number of people. In addition, recognizing that the current recording tools are stethoscopes and microphones which are not very convenient for portability, it would be more useful and practical if mobile devices with a built-in microphone can be used to perform the identification of breath sounds. In such a scenario, breath sounds acquired from mobile devices would be rather noisy and need to be enhanced.

REFERENCES

1. R. Darooei and A. Mahloojifar, "Asthma diagnosis based on respiratory dynamic using sparse representation based-classifier," in *Proceedings of the 22nd Iranian Con-*

- ference on Biomedical Engineering*, 2015, pp. 216-220.
2. E. Kaniusas, H. Pfützner, and B. Saletu, "Acoustical signal properties for cardiac/respiratory activity and apneas," *IEEE Transactions on Biomedical Engineering*, Vol. 52, 2005, pp. 1812-1822.
 3. R. L. Moedomo, M. S. Mardiyanto, and M. Ahmad, "The breath sound analysis for diseases diagnosis and stress measurement," in *Proceedings of International Conference on System Engineering and Technology*, 2012, pp. 1-6.
 4. Y. Ren, C. Wang, and J. Yang, "Fine-grained sleep monitoring: Hearing your breathing with smartphones," in *Proceedings of IEEE Conference on Computer Communications*, 2015, pp. 1194-1202.
 5. D. Chamberlain *et al.*, "Application of semi-supervised deep learning to lung sound analysis," in *Proceedings of Annual International Conference of IEEE Engineering in Medicine and Biology Society*, 2016, pp. 804-807.
 6. S. A. Taplidou and L. J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds," *Computers in Biology and Medicine*, Vol. 37, 2007, pp. 1073-1083.
 7. D. E. Olson and J. R. Hammersley, "Mechanisms of lung sound generation," *Seminars in Respiratory Medicine*, Vol. 6, 1985, pp. 171-179.
 8. R. G. Crystal *et al.*, "Chapter 100: Lung sounds," in *THE LUNG: Scientific Foundations*, 2nd ed., Philadelphia, USA, 1997.
 9. M. Sarkar, I. Madabhavi1, N. Niranjana, and M. Dogra, "Auscultation of the respiratory system," *Annals of Thoracic Medicine*, Vol. 10, 2015, pp. 158-168.
 10. P. Corbishley and E. Rodríguez-Villegas, "Breathing detection: Towards a miniaturized, wearable, battery-operated monitoring system," *IEEE Transactions on Biomedical Engineering*, Vol. 55, 2008, pp. 196-204.
 11. D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, Vol. 10, 2000, pp. 19-41.
 12. N. Dehak *et al.*, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 19, 2011, pp. 788-798.
 13. P. Matejka *et al.*, "Full-covariance ubm and heavy-tailed plda in *i*-vector speaker verification," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 4828-4831.
 14. J. M. K. Kua *et al.*, "*i*-vector with sparse representation classification for speaker verification," *Speech Communication*, Vol. 55, 2013, pp. 707-720.
 15. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, Vol. 2, 1998, pp. 121-167.
 16. V. N. Vapnik, *The Nature of Statistical Learning*, NY, Springer, 1995.
 17. C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, Vol. 13, 2002, pp. 415-425.
 18. V. N. Vapnik, *Statistical Learning Theory*, 1st ed., J. Wiley, N.Y., 1998.
 19. C. P. Chen and J. A. Bilmes, "MVA processing of speech features," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, 2007, pp. 257-270.
 20. M. C. A. Korba, D. Messadeg, H. Bourouba, and R. Djemili, "Noise robust features based on MVA post-processing," in *Proceedings of IFIP International Conference on Computer Science and its Applications*, Vol. 456, 2015, pp. 155-166.
 21. A. Lumb, "Chapter 1: Functional anatomy of the respiratory tract," in *Nunn's Applied Respiratory Physiology*, 8th ed., Elsevier, London, 2016, pp. 3-15.



Van-Thuan Tran received M.S. degree in Electrical Engineering and Computer Science from National Taipei University of Technology, Taipei, Taiwan, in 2018. He is currently pursuing the Ph.D. degree in Electronic Engineering at National Taipei University of Technology, Taiwan. From 2015 to 2016, he worked as an Engineer at JGCS Consortium, NSRP Project, Thanh Hoa, Vietnam. His research interests include multimedia signal processing and applied artificial intelligence.



Yung-Chih Lin (林勇志) received the M.S. degree in Computer and Communication Engineering from National Taipei University of Technology, Taiwan, in 2016. His research interests include signal processing and multimedia applications.



Wei-Ho Tsai (蔡偉和) received the Ph.D. degree in Communication Engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 2001. From 2001 to 2003, he was with Philips Research East Asia, Taipei, Taiwan, where he worked on speech processing problems in embedded systems. From 2003 to 2005, he served as a Postdoctoral Fellow at the Institute of Information Science, Academia Sinica, Taipei. He is currently a Professor in the Department of Electronic Engineering, National Taipei University of Technology, Taiwan. His research interests include spoken language processing and music information retrieval.