Distinction Method between Expiratory and Inspiratory Sounds Using Biological Sound Sensor

Shoya Makihira^{a,*}, Naoto Murakami^a, Tsunahiko Hirano^b, Keiko Doi^b, Kazuto Matsunaga^b
Seiji Nishifuji^a, Shota Nakashima^{a*}

^aFaculty of Engineering, Yamaguchi University, 2-16-1, Tokiwadai, Ube, Yamaguchi, 755-8611, Japan

^bFaculty of Medicine, Yamaguchi University, 1-1-1, Minamikogushi, Ube, Yamaguchi, 755-8505, Japan

*Corresponding Author: s-naka@yamaguchi-u.ac.jp

Abstract

The number of deaths from COPD in 2019 is about 6% of all deaths worldwide. The prevalence of COPD is expected to increase worldwide. Pulmonary function testing, so called spirometry is used in the diagnosis and severity assessment of COPD. There is a simple test method using expiratory and inspiratory time. Systems to measure expiratory and inspiratory time do not proposed. In this study, we propose a novel distinction method between expiratory and inspiratory sounds using biological sound sensors. The biological sound sensor consists of two units: holding and sensor units. The former fixes the sensor unit. The latter obtains biological sounds and adopts a polyurethane elastomer to match the acoustic impedance. The respiratory sounds are extracted by applying a bandpass filter to the biological sounds. Furthermore, Harmonic/Percussive Sound Separation is applied to the respiratory sounds to reduce the residual vascular sounds. The classifier between expiratory and inspiratory sounds is built with a soft margin Support Vector Machine. The feature is the power spectrum extracted from the spectrogram of respiratory sound. The classifier was built for each subject from the two respiration patterns. The proposed method was verified by the accuracy, precision, recall, and F-score. The obtained distinction accuracy was up to 86.8%, and it was possible to distinguish between expiratory and inspiratory sounds with high accuracy.

Keywords: Respiratory sound, Biological sound sensor, SVM, HPSS.

1. Introduction

Approximately 6% of the total 55.4 million deaths worldwide in 2019 were due to chronic obstructive pulmonary disease (COPD)⁽¹⁾. COPD is caused by inflammation of the lungs due to prolonged inhalation of toxins such as cigarette smoke, resulting in symptoms such as dyspnea on exertion⁽²⁾. The prevalence of COPD is expected to increase worldwide^(3, 4). Currently, Pulmonary function testing, so called spirometry is used in the diagnosis and severity assessment of COPD. For spirometry, an expensive device called spirometer is used. The spirometry is difficult to perform at non-medical institutions because it requires lung volume measurement. Thus, there is a simple test method using expiratory and inspiratory time. Expiratory time in COPD patients is extended over inspiratory time due to bronchial narrowing caused by airflow obstruction⁽⁵⁾. From this symptom, respiratory function status can be evaluated by the difference between expiratory and inspiratory time. This test method does not require expensive equipment such as the spirometer. It is also easy to diagnose non-medical institutions. In our laboratory, a sensor system for obtaining pulse and respiratory sound signal (RSS) using an electret condenser microphone was proposed^(6, 7). The system can obtain respiratory sounds non-medical institutions. In this study, we propose a system that uses machine learning to distinguish between exhalation and inhalation sounds in order to measure the time length of the two sounds. The respiratory sounds are obtained by the biological sound sensor. There is a difference in the frequency components of expiratory and inspiratory sounds in the human auditory range. For this reason, the power spectrum of expiratory and inspiratory sounds is used as a feature for machine learning.

This paper is organized as follows: In Chapter 1, the

background and motivation of the study are introduced. Chapter 2 describes the biological sound sensor to be used. Chapter 3 explains the distinction method between expiratory and inspiratory sounds using biological sound sensor. Chapter 4 describes the experimental methodology and evaluation functions. Chapter 5 concludes this study.

2. Biological Sound Sensor

Biological Sounds are generated from the body, and are transmitted through to the body surface. For example, vascular sounds, respiratory sounds, and swallowing sounds. In this chapter, respiratory sounds and biological sound sensors are described.

2.1 Respiratory Sound

The respiratory sounds are the sounds produced by the turbulence and eddy of gas flowing in the airways during external respiration⁽⁸⁾. External respiration refers to the exchange of gas between air and blood, by expansion and contraction of the lungs with the vertical movement of the diaphragm⁽⁹⁾. External respiration is of two kinds: inhalation and exhalation. Inhalation has the function of drawing air into the lungs and supplying oxygen to the blood vessels. Exhalation has the function of expelling carbon dioxide-rich air from the body. The frequency component of respiratory sounds, including inhalation and exhalation, is in the range of 200 ~ 2000 Hz. However, the high-frequency component decreases as it passes through the body. For example, the frequency component of respiratory sounds is about 200 ~ 700 Hz in the chest region⁽¹⁰⁾.

2.2 Biological Sound Sensor

In this study, we use a biological sound sensor to obtain the biological sounds. The sensor consists of two units: the sensor unit and the holding unit. The sensor unit is composed of an electret condenser microphone (EM-289 made by Primo Corporation, hereinafter called “ECM”) and a polyurethane elastomer (HITOHADA GEL hardness of Asker C 15 made by Exseal Corporation) in the light-curing resin case. The appearance and structure of the sensor unit is shown in Figs. 1a and 1b. The sensor unit adopts the ECM with an exposed diaphragm. The polyurethane elastomer serves to match the acoustic impedance between the sensor and the skin⁽¹¹⁾. It also serves to protect the ECM from oily stains on the skin surface. The light-curing resin case protects the sensor and holds the ECM. It also serves to reduce fricative noise and contact noise. The holding unit attaches

Fig. 1. Biological sound sensor.

from back of user’s head and serves to fix the sensor unit. The material of holding unit is elastic plastic. Fig. 1c shows the appearance of the holding unit with the sensor unit attached.

3. Proposed method

3.1 Outline

In this study, a distinction method between expiratory and inspiratory sounds is proposed based on the variation of power spectrum. First, the biological sound obtained by the sensor is applied a bandpass filter to extract the respiratory sound. Nevertheless, the noise of vascular sounds remains in the respiratory sounds after filtering. We use Harmonic/Percussive Sound Separation (HPSS) to remove the remaining noise in the respiratory sound. Machine learning is used to build a classifier from respiratory sounds with noise removed. The power spectrum of respiratory sounds is used as feature. Additionally, to prevent underfitting by individual differences in feature, classifiers are built for each subject.

3.2 Infinite Impulse Response (IIR) Bandpass Filter

Fig. 2 shows the biological sound obtained from the

Fig. 2. Example of the biological sound.

Fig. 4. Result after HPSS.

Fig. 3. Result after IIR bandpass filter.

sensor. The Biological sound include RSS and VSS. The frequency range of respiratory sounds through the human internal tissues is about 200 ~ 700 Hz. The frequency range of vascular sound signal (VSS) produced by the pulse is 75 ~ 200 Hz^(12, 13). The RSS is extracted by applying an IIR bandpass filter based on the frequency difference. The filter is designed to provide a sharp frequency response with low order. Fig. 3 shows the VSS extracted by the filter. The RSS is 60 dB smaller than the VSS, so the noise of VSS remains in the RSS after filtering. Thus, the filter only is not sufficient to extract the RSS.

3.3 Harmonic/Percussive Sound Separation (HPSS)

HPSS is a simple algorithm and effective method for separating the harmonic and percussive components of a monaural audio signals⁽¹⁴⁾. The harmonic components form horizontal ridges on the spectrogram. In contrast, the percussive components form vertical ridges of the broadband frequency response. In respiratory sounds, the harmonic component corresponds to RSS and the percussive

Fig. 5. Respiratory sound spectrogram.

component corresponds to VSS. This method applies median filtering to the spectrogram of an audio signal. To extract the RSS, a mask is created from the filtered spectrogram and applied to the original spectrogram. Fig. 4 shows the results of RSS obtained by using HPSS. The RSS is applied Short-Time Fourier Transform (STFT) to compute the spectrogram. The spectrogram of RSS is shown in Fig. 5. A spectrogram expresses an audio signal in terms of time, frequency, and power spectrum.

3.4 Support Vector Machine (SVM)

SVM is a supervised learning algorithm designed to solve binary classification problems⁽¹⁵⁾. The optimal boundary for classifying two classes is called the hyperplane. The goal of the binary classification problem is to find the hyperplane. The training data closest to the hyperplane is defined the support vector. In addition, the distance to the support vector is defined the margin. SVM has good generalization performance by determining the hyperplane with the largest margin. In this paper, expiratory and inspiratory sounds are regarded as two classes for machine learning. The training data is the power spectrum

corresponding to 0 ~ 1000 Hz from the spectrogram extracted at each time frame. The power spectrum indicates different frequency distributions for expiratory and inspiratory sounds, as shown in Fig. 5. Expiratory and inspiratory sounds have partially similar distribution and intensity in the power spectrum. Hence, the classifier is built using soft-margin SVM which can tolerate a little misclassification. Equations (1) and (2) show the objective function and constraints with the introduction of slack variables.

$$\min_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \right\} \quad (1)$$

$$t_i(\mathbf{w}^T \mathbf{x} + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (i = 1, 2, \dots, n) \quad (2)$$

Here, $\mathbf{w}$ is the normal vector of the hyperplane. $\mathbf{x}$ in the input vector. b is the linear coefficient of the hyperplane. ξ is a slack variable. C is the regularization parameter controlling the penalty for misclassification.

4. Experiment

4.1. Methodology

Biological sound was obtained by attaching the sensor unit to the mastoid process as shown in Fig. 6. The subjects were five males in their 20s who were instructed to remain seated and not to speak. The experiment conducted two one-minute biological sound measurements, each of two different respiratory patterns. Pattern I is normal respiratory, and pattern II assumed random breathing is two inhalations and two exhalations. The first biological sound of both patterns were used as training data to build the classifiers. The second two biological sounds were used as test data for classifier validation.

Respiratory sounds were extracted from biological sounds by the signal processing described above. Table 1 shows the experimental conditions. The IIR filter order was set to 12th order to reduce VSS included below 100 Hz. The obtained respiratory sounds were manually separated into expiratory and inspiratory sounds. The separated expiratory and inspiratory sounds were normalized and labeled. SVM was trained using a linear kernel with $C = 0.1$. The feature dimension was 46. To evaluate the classification accuracy, the classifier was applied to distinguish the test data. If more than 50% of the frames in the respiratory sound data were classified as exhalations, the classification result was defined as exhalation. Likewise, if the inhalations were more than

Fig. 6. Measurement conditions.

Table 1. Experimental conditions.

Parameters	
Sampling frequency	44.1 kHz
IIR filter response	Butterworth
IIR filter order	12
Window of STFT	Hanning
Frame length of STFT	2048 points
Shift length of STFT	256 points

50 %, the classification result was defined as inhalation. The classification results were compared with the correct labels.

The evaluation functions were accuracy, precision, recall, and F-score. The classification results of the learned model for a binary classification problem are represented as a confusion matrix. The confusion matrix is represented by four values, true positive (TP), true negative (TN), false positive (FP), and false negative (FN)⁽¹⁶⁾. In this experiment, we define exhalation as positive and inhalation as negative. FP is the inspiratory sound misclassified as the expiratory sound. FN is the inspiratory sound misclassified as the expiratory sound. The formulas for *Accuracy*, *Precision*, *Recall*, and *F-score* are shown in Equations (3) to (6). *Precision* and *Recall* are trade-offs. *F-score* is the harmonic mean of *Precision* and *Recall*.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F\text{-score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (6)$$

Precision is calculated from TP and FP . This evaluation function increases as the number of inspiratory

sound misclassifications decreases. The number of exhalations is the sum of *TP* and *FP*. Increased *TP* extends the expiratory time and risks being misdiagnosed as a decrease in respiratory function. The actual diagnosis is based on a respiratory function test followed identification of the disease by physicians. The false decrease in respiratory function is excluded by physicians. *Recall* is the percentage of correctly classified expiratory sounds out of the total expiratory sounds. *FN* is the value of a misclassified expiratory sound. Increased *FN* risks missing extended expiratory time.

4.2. Results and Discussion

The results for each subject in pattern I and pattern II as well as their weighted averages (WA) are shown in Tables 2 and 3. Table 4 shows WA of the two respiratory patterns for each subject. Subject A had about 80% *Accuracy* in both patterns I and II. WA of pattern I was 62.0%, almost the same value as pattern II. In pattern I for subject B, *Precision* was 83.3% and *Recall* was 88.2%. The results show that the classifier correctly distinguishes between expiratory and inspiratory sounds in normal respiration of subject B. On the other hand, *Accuracy* of subject B was 14.5% in pattern II. The *Precision* and *Recall* were similarly low. In contrast, in subjects C and E, respiratory pattern II was distinguished more accurately than pattern I. These results indicate that a classifier built from two respiratory patterns is at risk of specializing in distinguishing one respiratory pattern from the other. Two factors are considered to be the main causes of biased distinction. The first factor is that the power spectrum distribution of the respiratory sound is different between the two respiratory patterns. The second factor is the high dimension used in SVM. The hyperplane is determined using the features of the respiratory sound. Overfitting one respiratory pattern caused the other respiratory pattern to become outliers. As described in Chapter 3, expiratory and inspiratory sounds are partially similar in the power spectrum. In training with the high dimension, the hyperplanes were incorrectly determined by similar frequency bands of respirations. Therefore, high accuracy distinction can be expected in learning each respiratory patterns and by using new features that represent respiratory sounds.

5. Conclusions

This paper proposes a method for the distinction

Table 2. Result of Pattern I.

Subject	<i>Accuracy</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F-score</i> (%)
A	77.8	77.8	77.8	77.8
B	84.4	83.3	88.2	85.7
C	39.1	36.4	36.4	36.4
D	50.0	0	-	-
E	58.3	62.5	41.7	50.0
WA	62.0	50.0	63.3	55.9

Table 3. Result of Pattern II.

Subject	<i>Accuracy</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F-score</i> (%)
A	80.0	100.0	57.1	72.7
B	14.5	15.2	16.7	15.9
C	79.4	84.6	71.0	77.2
D	62.0	52.4	55.0	53.7
E	86.8	81.8	94.7	87.8
WA	60.5	58.2	56.1	57.1

Table 4. WA of the two respiratory patterns for each subject.

Subject	<i>Accuracy</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F-score</i> (%)
A	79.2	88.2	65.2	75.0
B	38.3	39.2	42.6	40.8
C	68.6	70.3	61.9	65.8
D	57.3	29.7	55.0	38.6
E	75.8	76.7	74.2	75.4
WA	61.0	55.2	58.3	56.7

between expiratory and inspiratory sounds using biological sound sensor. Machine learning such as SVM using the power spectrum of respiratory sounds as a feature. The proposed method can distinguish with a maximum accuracy of 86.8%. The proposed method is expected to be applied to a system that can evaluate respiratory function based on the time length of expiratory and inspiratory sounds from biological sounds.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP21K12797.

References

- (1) World Health Organization : “The top 10 causes of death”, Fact sheet, <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>, Date last updated : 9 December 2020, Date last accessed : 3 June 2021.
- (2) K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. V. Weel, J. Zielinski, Global Initiative for Chronic Obstructive Lung Disease : “Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease : GOLD Executive Summary”, *American Journal of Respiratory and Critical Care Medicine*, Vol. 187, Issue 4, pp. 347-365, 2013.
- (3) GBD 2015 Chronic Respiratory Disease Collaborators : “Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015”, *The Lancet Respiratory Medicine*, Vol. 5, Issue 9, pp. 691–706, 2017.
- (4) D. Adeloye, S. Chua, C. Lee, C. Basquill, A. Papana, E. Theodoratou, H. Nair, D. Gasevic, D. Sridhar, H. Campbell, K. Yee Chan, A. Sheikh, I. Rudan, Global Health Epidemiology Reference Group (GHERG) : “Global and regional estimates of COPD prevalence: Systematic review and meta-analysis”, *Journal of Global Health*, Vol. 5, No. 2, 2015.
- (5) P. Gruber, S. Swadron : “The Acute Presentation Of Chronic Obstructive Pulmonary Disease In The Emergency Department: A Challenging Oxymoron”, *Emergency Medicine Practice*, Vol. 10, No. 11, pp. 1-28, 2008.
- (6) K. Torii, H. Nakamura, T. Ooyagi, K. Tanaka, H. Yagami, S. Nakashima : “Distinction of Heart Sound and Respiratory Sound Using Body Conduction Sound Sensor Based on HPSS”, *ACIT2019*, pp. 198-201, 2019.
- (7) N. Murakami, K. Torii, S. Nakashima : “Biological Sound Sensor Robust to Air Conduction Noise”, *IEEE*, 2020.
- (8) R. Xaviero, A. Pramono, S. Bowyer, E. Rodriguez, “Automatic adventitious respiratory sound analysis: A systematic review”, *PLoS One*, Vol. 12, No. 5, pp. 1-43, 2017.
- (9) R. W. Flumerfelt, E. D. Crandall : “An analysis of external respiration in man”, *Mathematical Biosciences*, Vol. 3, pp. 205-230, 1968.
- (10) G. R. Wodicka, K. N. Stevens, H. L. Golub, E. G. Cravalho, D. C. Shannon : “A model of acoustic transmission in the respiratory system”, *IEEE Transactions on Biomedical Engineering*, Vol. 36, Issue 9, pp. 925-934, 1989.
- (11) A. I. Chen, M. L. Balter, M. I. Chen, D. Gross, S. K. Alam, T. J. Maguire, M. L. Yarmush : “Multilayered tissue mimicking skin and vessel phantoms with tunable mechanical, optical, and acoustic properties”, *Medical Physics*, Vol. 43, Issue 6, pp. 3117-3131, 2016.
- (12) M. J. Drinnnan, J. Allen, A. Murray : “Relation between heart rate and pulse transit time during paced respiration”, *Physiological Measurement*, Vol. 22, No. 3, p. 425, 2001.
- (13) H. V. Egidy : “The origin of the heart sounds”, *Basic research in cardiology*, pp. 395-441, 1973.
- (14) D. FitzGerald : “HARMONIC/PERCUSSIVE SEPARATION USING MEDIAN FILTERING”, *13th International Conference on Digital Audio Effects DAFx-10 Proceedings*, pp. 1-4, 2010.
- (15) N. Cristianini, J. S. Taylor : “An Introduction to Support Vector Machines and other kernel-based learning methods”, *CEMBRIDGE UNIVERSITY PRESS*, p. 7, 2000.
- (16) A. Baratloo, M. Hossein, A. Negida, G. E. Ashal : “Simple Definition and Calculation of Accuracy, Sensitivity and Specificity”, *Archives of Academic Emergency Medicine (Emergency)*, Vol. 3, No. 2, pp. 48-49, 2015.