CLASSIFICATION BETWEEN NORMAL AND ABNORMAL RESPIRATION USING ERGODIC HMM FOR INTERMITTENT ABNORMAL SOUNDS

Masaru Yamashita
Graduate School of Engineering
Nagasaki University
Nagasaki, Japan
masaru@cis.nagasaki-u.ac.jp

***Abstract*—Because there are many cases wherein abnormal sounds, which are called adventitious sounds, are included in the lung sounds of a subject suffering from pulmonary disease, the objective of this study was to automatically detect abnormal sounds from auscultatory sounds. To this end, we expressed the acoustic features of the normal lung sounds of healthy subjects and abnormal lung sounds of patients using hidden Markov models (HMMs), and distinguished between normal lung sounds and abnormal lung sounds. In our previous study, we constructed left-to-right HMMs with limited states. Because we expressed the abnormal sounds that occur intermittently and repeatedly using limited states, the HMMs could not express the acoustic features of abnormal sounds. Therefore, the classification rate between normal and abnormal respiration was low (86.53%). In this paper, we propose the construction of ergodic HMMs with a repetitive structure for intermittent abnormal sounds. By using a HMM that can express the acoustic features of abnormal sound in detail, the classification rate increased (88.81%). The results obtained by this study demonstrate the effectiveness of the proposed method.**

Keywords—hidden Markov model, lung sound, abnormal respiration, classification

I. INTRODUCTION

The auscultation of the lungs is a means of detecting patients suffering from pulmonary disease. Despite other non-invasive inexpensive methods, auscultation using a stethoscope can obtain valuable information regarding the health status of an individual. In many cases, abnormal sounds (called adventitious sounds [1]) are included in the lung sounds of a subject suffering from pulmonary disease, and auscultation is nowadays an effective method for diagnosing pulmonary disease. However, this method requires expert knowledge and experience. Therefore, identifying the difference between healthy and afflicted subjects is difficult for non-medical personnel, and this may be the reason that auscultation does not penetrate common households. Furthermore, it is difficult for the elderly or individuals living in depopulated areas to visit the hospital. Thus, the distinction between healthy and afflicted subjects at home will facilitate the early detection of pulmonary disease.

Several studies have focused on automatically detecting adventitious sounds from lung sounds [2–4]. These studies either detected a specific adventitious sound using a wavelet transform or distinguished the frame of an adventitious sound using the short-time spectrum. However, the time of occurrence and the duration of adventitious sounds vary. Therefore, it is desirable to discriminate sounds using the features of the entire respiration and its inflection. Furthermore, the features of adventitious and respiratory sounds depend on each individual and the progress of the disease. Therefore, we considered that these features should be expressed statistically. In a previous study, we expressed the time-series of the acoustic features of lung sound by constructing hidden Markov models (HMMs), and discriminated between normal and abnormal respiratory sounds [5, 6]. Moreover, we constructed HMMs with high accuracy by selecting a suitable number of states and mixtures for adventitious sound segments [7]. However, we did not consider the suitable state transition of the HMMs. Fig. 1 shows an example of respiratory sounds that include adventitious sounds called fine crackle. Adventitious sounds are divided into two classes: continuous adventitious sounds and discontinuous adventitious sounds. For example, fine crackle is a type of discontinuous adventitious sound.

The distinctive feature of discontinuous adventitious sounds is that short sounds occur repeatedly. We considered the discontinuous adventitious sound period as a steady state and expressed it using a left-to-right HMM. However, the classification rate between normal and abnormal respiration was low. Hence, we concluded that these models are not suitable and focused on analyzing the topology of acoustic models.

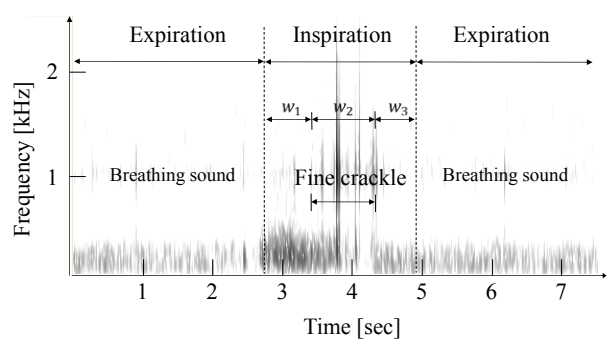

Fig. 1. Example of respiratory sound including adventitious sound.

In this paper, we propose the construction of an ergodic HMM for discontinuous adventitious sounds. To construct the ergodic HMM, we set a suitable analysis frame length and appropriate frame intervals.

II. Lung Sound Database

A. Hand Labeling

We recorded lung sounds using an electronic stethoscope. Subsequently, we performed segmentation manually based on the recorded sounds, waveform, spectrogram, and power. First, we divided the lung sounds into inspiration and expiration sound segments (respiratory sound segments). Next, we divided the respiratory sound segments into continuous adventitious sound segments, discontinuous sound segments, and other breathing sound segments. If the occurrence interval of the adventitious sounds was shorter than 100 ms, we considered this as a single segment.

B. Definition of Normal and Abnormal Respiration

The acoustic features of some noises are similar to those of adventitious sounds. Some respiratory sounds produced by healthy subjects include noises, which makes diagnosis difficult for individuals without medical training. Conversely, some respiratory sounds produced by patients do not include adventitious sounds. However, we cannot term these sounds as normal respiratory sounds. Therefore, normal and abnormal respiration must be defined. In this study, respiratory sounds were grouped into four categories.

-Abnormal respirations of patients (AP): the respirations include adventitious sounds produced by patients.

-Abnormal respirations of healthy subjects (AH): the respirations include noises resembling the adventitious sounds produced by healthy subjects.

-Normal respirations of patients (NP): the respirations do not include adventitious sounds or noises resembling the adventitious sounds produced by patients.

-Normal respirations of healthy subjects (NH): respirations do not include adventitious sounds or noises resembling the adventitious sounds produced by patients.

In the experiment for discriminating between normal and abnormal respiration, we considered NH as normal respiration and AP as abnormal respiration. In other words, we did not use AH and NP. However, we did consider all respirations in the experiment for discriminating between healthy subjects and patients.

III. Classification Procedure

A. Procedure for Classification between Normal and Abnormal Respiration

Generally, in the field of speech recognition, the acoustic models of the phoneme (as the smallest unit of speech) and the occurrence probability of words are used to construct stochastic models. In this study, this technique was applied to lung sounds. Fig. 2 shows the architecture of the system for classifying normal and abnormal respiration [6].

The classification procedure consists of training and test processes. In the training process, the HMMs, as the acoustic model and segment sequence model [6] that defines the occurrence probability of the divided segments, are trained. In the test process, the input respiration is distinguished as normal or abnormal respiration based on the maximum likelihood approach. If we assume that the sample respiration W consists of N segments, it can be expressed as $W = w_1 w_2 \cdots w_i \cdots w_N$, where w_i is the ith segment of W.

The training process is as follows. First, we extract the acoustic features and train each segment. In the case of normal respiration, there is one segment (N=1). Conversely, in the case of abnormal respiration, which includes adventitious sounds, there are at least two segments ($N \geq 2$). For example, the inspiration case shown in Fig. 1 consists of one fine crackle segment and two breathing segments (N=3). The expiration case shown in Fig. 1 does not include adventitious sounds, and consists of one breathing sound segment (N=1). The training of the segment sequence model is as follows. The occurrence probability of segments $P(W)$ is calculated using a segment bigram; $P(W)$ can be written as follows:

$$P(W) = w_1 \times \prod_{i=2}^{N} P(w_i | w_{i-1}). \quad (1)$$

Let $P(w_i|w_{i-1})$ be defined as

$$P(w_i|w_{i-1}) = C(w_i|w_{i-1}) = (w_{i-1}, w_i)/C(w_{i-1}), \quad (2)$$

where $C(w_i)$ is the count of w_i, $C(w_{i-1})$ is the count of w_{i-1}, and $C(w_i|w_{i-1})$ is the count of segment w_i after w_{i-1} in the training database.

The test process is as follows. The maximum likelihood among the calculated likelihoods is determined, and the corresponding segment sequence $\widehat{W}$ is selected to identify the

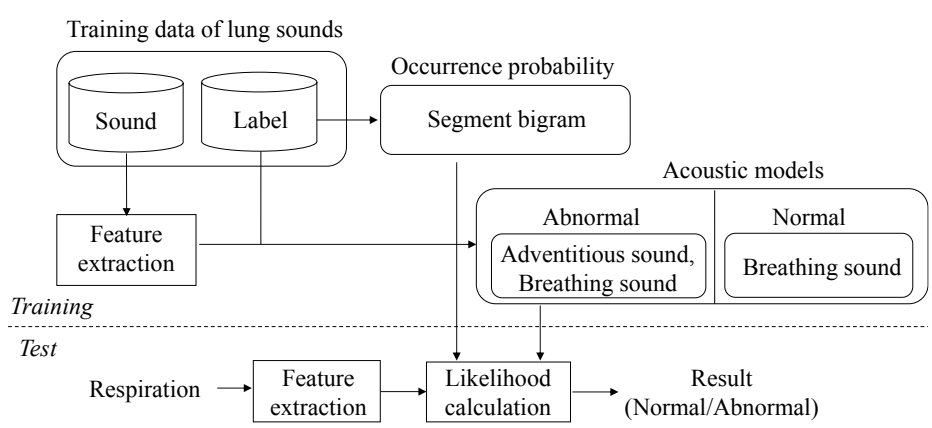

Fig. 2. Architecture of system for classification between normal and abnormal respiration.

sample respiration sound. If the sequence includes at least one adventitious sound, we identify the sample respiration as an abnormal sound. Otherwise, we identify the sample respiration as a normal sound.

$\widehat{W}$ can be written as follows:

$$\widehat{W} = \underset{W}{\operatorname{argmax}}(\log P(X|W) + \alpha \log P(W)), \quad (3)$$

where X is the sample respiration and $\log P(X|W)$ is the acoustic likelihood. The weight factor was obtained experimentally.

B. Procedure for Classification between Healthy Subjects and Patients

This section describes the detection of patients by using a series of respirations [8]. Noises from outside of the body occur irregularly. In contrast, adventitious sounds occur periodically. Therefore, in the case of healthy subjects, most respirations are classified as normal, even if one or a few respirations are classified as abnormal respiration. In other words, for healthy subjects, most of the likelihood values for normal respiration are higher than the likelihood values for abnormal respiration, even if one or a few respirations are classified as abnormal respiration. For the detection of patients, we calculate the likelihood $\mathrm{L}(W_{no})$ for the segment sequence W_{no} that does not include adventitious sounds and the maximum likelihood $\mathrm{L}(W_{ab})$ for the segment sequence W_{ab} that includes adventitious sound segments, for each respiration. If the total $\mathrm{L}(W_{ab})$ is greater than or equal to the total $\mathrm{L}(W_{no})$, the subject is classified as a patient. In other words, the following relationship holds:

$$\sum_j \mathrm{L}(W_{j,ab}) \geq \mathrm{L}(W_{j,no}) \quad (4)$$

where $\mathrm{L}(W_{j,ab})$ is the likelihood of the segment sequence that includes adventitious sound segments in the jth respiration of the subject, and $\mathrm{L}(W_{j,no})$ is the likelihood for the segment sequence that does not include adventitious sound segments in the jth respiration of the subject. For each auscultation point, we classify the respiration series as a healthy subject or a patient.

IV. Construction of Appropriate HMM for Discontenious Adventitious Sounds

In our previous studies [5-8], we used left-to-right HMMs for each segment, as shown in Fig. 3(a), and concluded that these models are not suitable. Therefore, we focus on analyzing the topology of acoustic models. To construct HMM that is appropriate for discontinuous adventitious sounds, we assume that the discontinuous sound period consists of the repetition of an abnormal sound period and a silent period. Then, we construct the ergodic HMM as shown in Fig. 3(b). Furthermore, to construct the ergodic HMM, we select the suitable analysis frame length and frame intervals because an abnormal sound period and a silent period are too short to analyze using typical values used in the frequency analysis of speech.

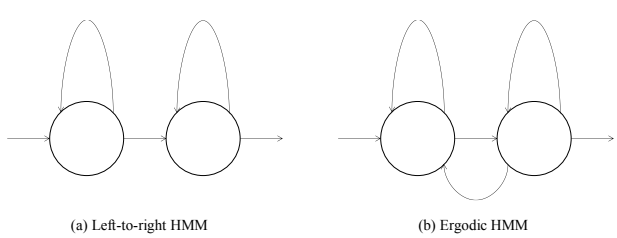

Fig. 3. Topology of HMMs.

V. Evaluation Experiments

A. Experimental Conditions

Six Mel-Frequency Cepstrum Coefficients (MFCCs) and the power were extracted as acoustic features using a Hamming window. The lung sound data were sampled at 5 kHz. Fig. 4 shows the auscultation points. In this study, the auscultated lung sounds from nine points (P_1 - P_9) were considered in the experiments. The number of abnormal respiratory sounds and number of patients are listed in TABLE I. As many normal respirations and healthy subjects were randomly selected for the experiments. We performed leave-one-out cross-validation to construct a subject-independent model.

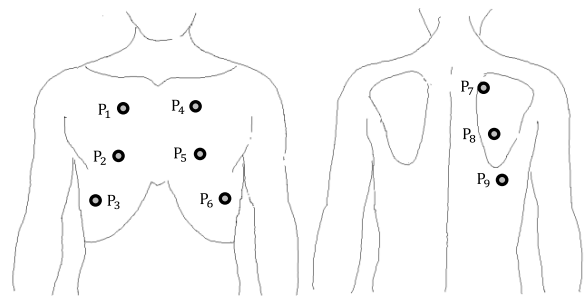

Fig. 4. Auscultation points.

TABLE I. Number of Abnormal Respiratory Sounds and Patients

Auscultation Points	Number of Abnormal Respirations	Number of Patients
P_1	219	44
P_2	161	89
P_3	254	53
P_4	217	47
P_5	312	62
P_6	206	52
P_7	182	46
P_8	324	62
P_9	260	62
Total	2135	517

B. Classification Experiments

We compared the left-to-right HMM and ergodic HMM and selected the analysis frame length and frame intervals. First, we set the analysis frame length to 25 ms and the frame interval to 10 ms, as in our previous study [5–8]. Subsequently, we selected several frame length and frame interval combinations, as presented in TABLE II. Fig. 5 shows the classification rate between normal and abnormal respiration using left-to-right HMM. When the analysis frame length and frame interval were too small, the classification rate decreased because the frequency resolution was low. Fig. 6 shows the classification rate between normal and abnormal respiration using ergodic HMM. The accuracy was lower than that of the left-to-right HMM for which the analysis frame length was set

to 25 ms and the frame interval was set to 10 ms. The reason for this is that the analysis frame length and frame interval were too large to express the acoustic features of each intermittent sound. Then, we set the frame length and frame interval to a somewhat small value (Condition C) and the classification accuracy increased. This result indicates the significant effectiveness ($p = 0.0007$) of constructing an ergodic HMM of discontinuous adventitious sounds. When a suitable analysis frame length and an appropriate frame interval were set, the ergodic HMM could express the acoustic features of intermittent sounds.

TABLE II. COMBINATIONS OF FRAME LENGTH AND FRAME INTERVAL

Conditions	Frame Length	Frame Interval
A	5	2
B	10	4
C	15	6
D	20	8
E	25	10
F	30	12

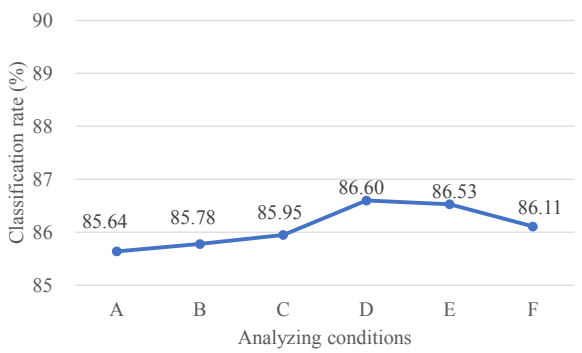

Fig. 5. Classification rate between normal and abnormal respiration using left-to-right HMM.

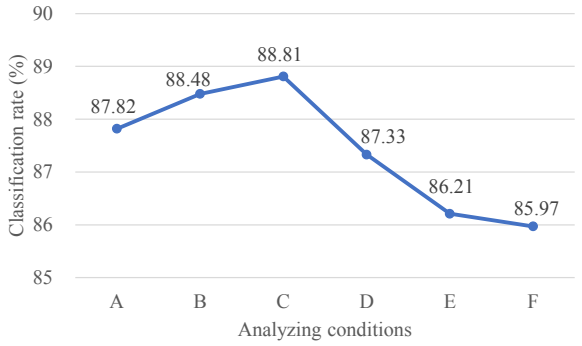

Fig. 6. Classification rate between normal and abnormal respiration using ergodic HMM.

Finally, the experiment for classification between healthy subjects and patients is discussed. TABLE III presents the classification rate of healthy subjects and patients. Both models use the best combination identified in the previously mentioned experiments to classify normal respiration and abnormal respiration.

The results obtained by the experiment for classifying healthy subjects and patients indicate that there is room for improving the classification rate. In this study, however, the improvement was not significant because the number of test subjects was small.

TABLE III. CLASSIFICATION RATE BETWEEN HEALTHY SUBJECT AND PATIENT (%)

Type of HMM	Healthy Subject	Patient	Average
Left-to-right HMM	88.8	84.4	86.4
Ergodic HMM	90.4	86.2	88.3

VI. CONCLUSIONS

This paper proposes the construction of an ergodic HMM for abnormal sounds that occur intermittently and repeatedly. To construct the HMM with a repetitive structure, we set a suitable analysis frame length and appropriate frame intervals. The results obtained by the classification experiment confirm that the classification rate improved when the frame length and frame interval were set to be slightly smaller than the typical values used in the frequency analysis of speech. Thus, the effectiveness of the proposed approach is demonstrated. However, in the experiment for classifying healthy subjects and patients, the improvement was not significant because the number of test subjects was small.

In future work, we will clarify the suitable topology of HMMs using a deep neural network, which has been proven to be effective in the field of speech recognition.

ACKNOWLEDGMENT

This study was supported by the Kawai Foundation for Technology & Music.

REFERENCES

[1] N. Gavriely and David W. Cugell, Breath sounds methodology. CRC Press, 1995.

[2] Y. P. Kahya, S. Yerer, and O. Cerid, "A wavelet-based instrument for detection of crackles in pulmonary sounds," EMBC 2001 (International Conference of the IEEE Engineering in Medicine and Biology Society), 2001.

[3] A. Marshall and S. Boussakta, "Signal analysis of medical acoustic sounds with applications to chest medicine," Journal of the Franklin Institute vol. 344, pp. 230–242, May 2007.

[4] S. A. Taplidou and L. J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds," Computers in Biology and Medicine, vol. 37, pp.1073–1083, August 2007.

[5] S. Matsunaga, K. Yamauchi, M. Yamashita, and S. Miyahara, "Classification between normal and abnormal respiratory sounds based on maximum likelihood approach," Proc. of ICASSP 2009 (IEEE International Conference on Acoustics, Speech, and Signal Processing), pp. 517–520, April 2009.

[6] H. Yamamoto, S. Matsunaga, M. Yamashita, K. Yamauchi, and S. Miyahara, "Classification between normal and abnormal respiratory sounds based on stochastic approach," Proc. of ICA 2010 (International Congress on Acoustics), Paper No. 671, August 2010.

[7] M. Yamashita, "Constructiion of effective HMMs for classification between normal and abnormal respiration," Proc. of APSIPA 2020, pp. 914–918, December 2020.

[8] S. Matsutake, M. Yamashita, and S. Matsunaga, "Discrimination between healthy subjects and patients using lung sounds from multiple auscultation points" Proc. of ICASSP 2013 (IEEE International Conference on Acoustics, Speech, and Signal Processing), pp. 1296–1300, May 2013.