Training Data Expansion for Classification between Normal and Abnormal Lung Sounds

Naoki Umeno, Masaru Yamashita, Hiroyuki Takada and Shoichi Matsunaga
Nagasaki University, Nagasaki, Japan
E-mail: {b119306, masaru, htakada, mat} @cis.nagasaki-u.ac.jp Tel: +81-95-8192700

Abstract— **In this paper, we investigate the effectiveness of training data expansion methods to distinguish between normal and abnormal lung sounds. Acoustic characteristics of lung sounds vary according to auscultation points. In conventional classification methods, acoustic models were usually trained using only lung sounds recorded at the same auscultation points to that of evaluation data. This results in a small amount of training data and, thus, hinders the achievement of a high classification rate. To overcome this problem, we performed training data expansion by selecting the lung sounds, which are expected to be useful for generating acoustic models with higher classification performance, among sound samples recorded at other auscultation points. We investigated the two types of selection approach: selection based on the similarity of acoustic features in sound samples and selection based on the confidence measure represented by the difference between the acoustic likelihood for a normal or abnormal respiratory candidate. Our experiments showed that both selection types have the potential to increase the classification performance between normal and abnormal lung sounds, as well as the classification performance between healthy and unhealthy subjects.**

I. INTRODUCTION

Auscultation is one of the most popular and cost-effective medical examination methods for identifying respiratory illnesses. This is because lung sounds of individuals with respiratory disorders frequently contain abnormal respiratory sounds known as adventitious sounds [1]. There are several types of adventitious sounds that depend on the condition of the abnormal internal organs and the degree of inflammation. However, it is difficult for individuals without expertise in auscultation to correctly identify types, for an accurate diagnosis. For this reason, automated determination of respiratory diseases, using respiratory sounds, is required.

Several studies to detect specific adventitious sounds have been performed based on the acoustic analysis of respiratory sounds [2-7]. The primary purpose of these studies was to assist doctors in making diagnoses. We have developed a classification procedure for distinguishing between normal and abnormal respiration, based on a maximum likelihood (ML) approach, using hidden Markov models (HMMs) [8-10]. Our purpose is to easily detect unhealthy subjects at home. In our detection procedure, the acoustic likelihoods for an abnormal and a normal respiration candidate were compared for the classification. This procedure demonstrated the usefulness of a stochastic approach in the detection of abnormal respiratory sounds in unhealthy subjects. However, we generate stochastic acoustic models by using a limited amount of training data that has been recorded from the same auscultation points to that of the evaluation sample. This is because of the difference in acoustic characteristics, caused by the acoustical variety of adventitious sounds and frequent contamination into lung sounds by noises. These phenomena depend on auscultation points. For example, in unhealthy subjects, lung sounds recorded from upper auscultation-points contain continuous adventitious lung sounds, such as wheezes and rhonchi, while sounds recorded at lower points contain discontinuous adventitious sounds, such as fine and coarse crackles. Heart sounds appear in the sounds recorded at the front left [10]. Then, when lung sounds recorded at different auscultation points from the evaluation sample were used in the training of acoustic models, classification performance usually decreased.

To address this problem, we investigated expansion methods that could increase the amount of training data. We assumed that all lung sound data recorded at different auscultation points (in reference to recordings at a specific point) were not useful for generating acoustic models for the target sample classification. Even though some samples were recorded at other points, and their acoustic characteristics were similar to those of data recorded at the target point, these data could be useful for generating the point-dependent acoustic models. Therefore, in our methods, we have selected the samples among those recorded at other points, then, both all samples recorded at the target point and the selected samples were used to generate topic-dependent acoustic models. In this paper, we investigated two methods for expansion. One is based on the similarity of distribution in the acoustic feature parameters between the target sample and samples recorded at a specific point. The other is based on the confidence of the target sample classification result [9], which was calculated from the difference between acoustic likelihood for a normal and abnormal respiratory candidate. Even though these fundamental techniques to measure the similarity have been used in various scientific areas, the

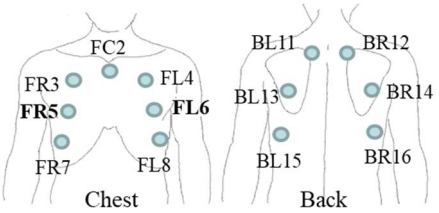

Fig. 1 Auscultation points on chest and back.

selection of lung sounds recorded at different auscultation points used to generate better acoustic models is an original approach. In this paper, we indicate the effectiveness of these two methods experimentally, discussing their advantages and disadvantages.

II. LUNG SOUND DATA

A. Training and Evaluation Data

Using an electronic stethoscope, which incorporated a piezoelectric microphone, we recorded lung sounds at thirteen typical auscultation points in healthy and unhealthy subjects with pulmonary emphysema, indicating a sign for the beginning inspiration and expiration phase to the subjects. The auscultation points are shown in Figure 1, where FR5 and FL6 are second intercostal spaces. At most one lung sound sample for each auscultation point, in each subject, was recorded. Lung sound samples from each subject were recorded one at a time. The total number of recorded samples from unhealthy subjects was 605 and that from healthy subjects was 837. The average number of respiratory phases per sample was approximately 10.

B. Manual Segmentation of Lung Sound

A lung sound sample S comprised several successive respiratory phases X:

$$S = X_1 \cdots X_l \cdots X_L, \quad (l = 1, \cdots, L) \tag{1}$$

where X_l is the l-th phase in which each respiratory period was manually detected. In our expansion approach, we performed a respiratory-phase-based (X_l) expansion and sample-based (S) expansion.

We prepared labels corresponding to the acoustic segments w based on acoustic [1] and temporal features manually [9]. In our labeling, we assumed that an abnormal respiratory phase was composed of N successive acoustic segments and that a normal respiratory phase comprised one breath segment. Then, a segment (symbol) sequence W of a respiratory phase (period) X is represented as

$$W = w_1 \cdots w_i \cdots w_N, \quad (i = 1, \cdots, N), \tag{2}$$

where the i-th acoustic segment is given as w_i. In our research, each adventitious sound was presented using a continuous acoustic or a discontinuous acoustic segment. These segments, and their boundaries, were used in a training process described in section III.

III. DATA EXPANSION STRATEGY

A. Flow of Data Expansion and Classification

The architecture of our classification system between normal and abnormal respiration, including our data expansion procedure, is shown in Figure 2. The system comprises two processes: training and testing.

In the training process, three types of procedure are used to generate acoustic models. One procedure is to generate conventional acoustic HMMs for each kind of segment and respiratory phase (inspiration/expiration). Because of the specific acoustic characteristics for each auscultation point, the conventional HMMs were generated using lung sounds recorded from the same auscultation points. The classification method that uses these models is referred to as Baseline I, in this paper. The other two procedures used a different data expansion function. These are described in the following subsection.

In the test process, for a respiratory phase X in a test sample, the likelihoods for a normal and abnormal candidate were calculated for each respiration phase, using each type of HMMs. The likelihood is composed of the acoustic likelihood calculated from HMMs θ and the segment-sequence likelihood $\log P(W)$, derived from segmental bigram, where P indicates probability. The segment (sequence) $\hat{W}$ with the highest likelihood for the respiratory phase is given using Bayes' theorem:

$$\begin{aligned}\hat{W} &= \underset{W}{\operatorname{argmax}} P(W|X,\theta) \\ &\approx \underset{W}{\operatorname{argmax}} (\log P(X|W,\theta) + \gamma \log P(W))\end{aligned} \tag{3}$$

The weight factor γ was obtained experimentally to achieve higher performance. Then, the most likely segment sequence $\hat{W}^{ab}$ for an abnormal respiratory candidate and its likelihood $\log P(\hat{W}^{ab}|X,\theta)$; and the likelihood $\log P(W^{no}|X,\theta)$ for a normal candidate W^{no} were obtained. The classification result was obtained by comparing these likelihoods.

B. Lung Sound Data Expansion

Basic strategy for data expansion is to select additional training samples from the set of lung sounds recorded at other

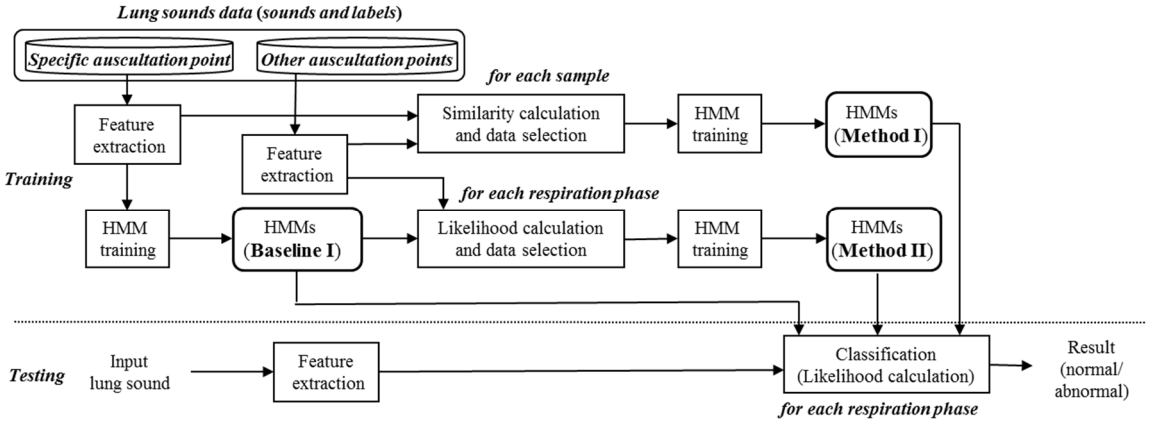

Fig. 2 Flow of data expansion and classification between normal and abnormal respirations.

auscultation points. We assumed that acoustic characteristics of effective samples from other points were similar to those of samples recorded at the target point. Based on this assumption, we devised two methods: one (Method I) based on the similarity between the acoustic characteristics distribution of each candidate sample, and that of all samples recorded at target point; and the other (Method II) based on the confidence of each candidate sample classification result. This confidence was calculated through the difference between the acoustic likelihood for a normal and abnormal respiratory candidate. Although both methods took into account the similarity of acoustic characteristics, Method I dealt with the similarity more directly.

1) Method I: Each candidate sample *S*, used for data expansion, is examined to check if it can be used as an additional training sample, based on the similarity in distributions of acoustic feature parameters (in this study, mel-frequency cepstral coefficients (MFCC) and power were used). To calculate the similarity, we used Bhattacharyya distance between the distribution of the samples recorded at the target auscultation-point, and that of the candidate. A smaller distance indicates similar acoustic characteristics.

First, we calculated three average distance values ($I_{NH}, I_{NU},$ and I_{AU}) using all samples recorded at same auscultation points. I_{NH} was calculated using normal lung sounds recorded from healthy subjects. While using the samples recorded from unhealthy subjects, I_{NU} was obtained using normal phases not including adventitious sounds, and I_{AU} was calculated using abnormal phases including adventitious sound segments. Next, labelled for the same conditions as above, three distance values ($O_{NH}, O_{NU},$ and O_{AU}) between the distribution of all samples recorded at the target point and that of each candidate sample recorded at other points.

For a lung sound candidate recorded from a healthy subject, if $O_{NH} < \alpha_{NH} I_{NH}$, this candidate was selected as a training sample. For candidate recording from an unhealthy subject, if both $O_{NU} < \alpha_{NU} I_{NU}$ and $O_{AU} < \alpha_{AU} I_{AU}$ were satisfied, this candidate was selected. Here, α is a constant which controls the amount of training data. For small α, a higher similarity was required to satisfy the conditions. Consequently, the training data stayed relatively small. This selection was performed for each lung sound sample *S*.

2) Method II: In each candidate phase X, we calculated likelihood difference d between the likelihood of normal phase $\log P(W^{no}|X,\theta)$ and that of abnormal phase $\log P(\widehat{W}^{ab}|X,\theta)$, where θ indicates the HMMs generated using lung sounds recorded at a specific point. If *X* is a phase from a normal subject, then *d* is represented as follows:

$$d = \log P(W^{no}|X,\theta) - \log P(\widehat{W}^{ab}|X,\theta) \quad (4)$$

Otherwise, if *X* is a phase from an unhealthy subject, we set $d = -d$. A candidate was selected as an additional training sample when d was larger than the predefined threshold D_T. If D_T was large, acceptance as training data required samples with high confidence. If D_T was zero, the samples that were recognized correctly by using HMMs θ were selected. Selection in this method was performed for each phase *X*.

For the proposed methods, both the samples recorded at a specific auscultation points and the selected samples, were used to generate new HMMs.

IV. Evaluation Experiments

A. Experimental Conditions

We performed classification tests to evaluate the effect of expanding training lung-sound samples. The lung sound data were sampled at 5 kHz. Every 10 ms, a vector of 6 MFCCs and power was computed using a 25-ms Hamming window. The acoustic HMMs for normal respiration were generated using the respiratory sounds from healthy subjects, and the models for adventitious sound segments were generated using the sounds from unhealthy subjects. HMMs with three states and two Gaussian probability density functions were used.

For the respiratory phase test samples there were 254 normal phases from 53 healthy subjects and 254 abnormal phases from 53 unhealthy subjects, recorded at auscultation points FR5, and 217 normal phases from 47 healthy subjects and 217 abnormal phases from 47 unhealthy subjects, recorded at FL6 in Figure 1. These samples were used for the training of the auscultation-point-dependent original HMMs. However, we performed a leave-one-out cross validation for each test phase. The numbers of lung sound samples for data expansion were described in Section II.

B. Classification of Normal and Abnormal Respiration

First, to investigate the variation of acoustic characteristics we calculated the three average values ($I_{NH}, I_{NU},$ and I_{AU}) of Bhattacharyya distance using samples recorded at FR5. Table I shows each average value and standard deviation. This table indicates that the variation in acoustic characteristics of abnormal phases is larger than that of normal phases. Variation from unhealthy subjects is also larger than that from healthy subjects.

Next, we prepared two types of baseline HMMs. One (Baseline I) was generated using the lung sounds recorded at the same auscultation point to that of the test input phase. The other was auscultation-independent HMMs (Baseline II) generated by using all samples. We also prepared two other types of HMMs by performing our data expansion based on the procedure of Methods I and II. We then carried out classification experiments, using each type of the HMMs, to distinguish between normal and abnormal respiratory phases. Classification performance for each method, and the approximate number of additional respiratory phases, are

TABLE I
AVERAGE DISTANCE VALUES BETWEEN LUNG SOUNDS RECORDED AT SAME AUSCULTATION POINT FR5 [%]

Phase			Average	S. D.
Type	Respiratory	Subject		
I_{NH}	Normal	Healthy	0.76	0.69
I_{NU}	Normal	Unhealthy	1.38	1.30
I_{AU}	Abnormal	Unhealthy	2.55	2.45

shown in Table II. In this table, the best performances from using Method I and II are listed, which were obtained by varying thresholds α and D_T. The average classification rate of 87.4% for Baseline II was lower than that of Baseline I: 89.8% ($p = 0.048$). This result indicates that exhaustive expansion of lung sound samples recorded at other auscultation points, is not necessarily tied to the improvement of classification performance. Meanwhile, both the classification rate of 91.5% for Method I and 91.4% for Method II were higher than those of Baselines I and II ($p = 0.10$ and 0.12, respectively). This validates training sample expansion by selecting samples based on acoustic similarity.

Finally, we compare Method I with Method II. For test samples recorded at FR5, Method I achieved a rate of 91.1 % by using 553 additional phases, while Method II achieved 90.7 % by using 518 additional phases; this demonstrated that Method I has a better potential than Method II. However, the performance of Method I decreased drastically to 88.6 % when the number of additional samples was 1320, whereas Method II maintained performance at 90.6 % when all samples classified correctly were used ($D_T = 0$); the total number of additional samples was approximately 4700. These results show that the performance of Method II is robust to the fluctuation of thresholds values, for the selection of additional samples. Because it is difficult to determine the most appropriate thresholds in advance, we conclude that Method II is more practical than Method I.

C. *Classification of Unhealthy and Healthy Subjects*

We performed experiments using the aforementioned four classification methods on unhealthy and normal subjects. Classification results were obtained by comparing the sum of likelihoods for all normal respiratory candidates $\sum_{l=1}^{L} \log P(W^{no}|X_l, \theta)$, with that of all abnormal respiratory candidates $\sum_{l=1}^{L} \log P(\widehat{W}^{ab}|X_l, \theta)$ in a test sample *S*.

The obtained classification rates are shown in Table III. According to the small number (200) of evaluation samples, sufficient analysis is difficult. However, the table indicates the proposed methods show promise for the classification between healthy and unhealthy subjects.

V. CONCLUSIONS

This paper presents the expansion methods of training data to distinguish between normal and abnormal lung sounds, by performing the selection of suitable lung sound samples from those recorded at different auscultation points. We investigated two expansion methods based on the similarity of acoustic characteristics between the samples recorded at target point and those recorded at other points. One selection method was based on the similarity of distribution of acoustic features and the other was based on the difference of likelihoods between the likelihood of the correctly classified normal/abnormal candidate and that of the other abnormal/normal candidate. According to the experimental results obtained for the classification between normal and abnormal respirations and those for the classification between healthy and unhealthy subjects, the performance of two proposed methods, which adopted the data selection based on the acoustical similarity, were better than that of the baseline without data expansion. Even though the selection method based on the similarity of distribution of acoustic features achieved higher performance, from the view point of robustness to set predefined thresholds, we think the selection method using the likelihood difference is more practical. However, we need to devise an appropriate method to set the most suitable thresholds for practical use of data expansion in future.

TABLE II
CLASSIFICATION PERFORMANCE BETWEEN NORMAL AND ABNORMAL LUNG SOUNDS AND NUMBER OF ADDITIONAL PHASES

Method	Normal [%]	Abnormal [%]	Average [%]	Additional phases
Baseline I	90.4	89.2	89.8	0
Baseline II	84.7	90.0	87.4	5460
Method I	91.1	91.9	91.5	550
Method II	90.0	92.8	91.4	520 ~ 4720

TABLE III
CLASSIFICATION PERFORMANCE BETWEEN HEALTHY AND UNHEALTHY SUBJECTS [%]

Method	Healthy	Unhealthy	Average
Baseline I	84	89	88.5
Baseline II	86	88	87.0
Method I	95	84	90.0
Method II	87	91	89.0

REFERENCES

[1] Noam Gavriely, "Breath sounds methodology," CRC Press, 1995.

[2] Y. P. Kahya, S. Yere and O. Cerid, "A wavelet-based instrument for detection of crackles in pulmonary sounds," *Proc. of IEEE EMBS*, pp.3175-3178, 2001.

[3] M. Bahoura and X. Lu, "Separation of crackles from vesicular sounds using wavelet packet transform," *Proc. of IEEE ICASSP*, II, pp. 1076-1079, 2006.

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

[5] 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, 2007.

[6] C. Daniel, et al., "Application of semi-supervised deep learning to lung sound analysis," *Proc. of IEEE EMBC,* WeCT 1.32, 2016.

[7] M. Aykanat, O. Kilig, B. Kurt and S. Saryal. "Classification of lung sounds using convolutional neural networks," *EURASIP Journal on Image and Video Processing,* 2017

[8] S. Matsunaga, et al., "Classification between normal and abnormal respiratory sounds based on maximum likelihood approach," *Proc. of IEEE ICASSP*, pp. 517-520, 2009.

[9] M. Yamashita, et al., "Discrimination between healthy subjects and patients with pulmonary emphysema by detection of abnormal respiration," *Proc. of ICASSP,* pp. 693-696, 2011.

[10] M. Yamashita, et al., "Robust classification between normal and abnormal lung sounds using adventitious-sound and heart-sound models," *Proc. of IEEE ICASSP*, pp. 4451-4455, 2014.