Analyzable Breath and Heart Sound Extraction for Cardiovascular Disease Monitoring

Lurui Wang, Zhongwei Jiang
Yamaguchi University, Yamaguchi, Japan

Abstract: Many studies have shown that sleep-related disorders cause a high risk of cardiovascular disease(CVD). Estimation of the CVD risk by analysis of the heart and breath sound is considered a convenient way in the research. An overnight heart sound and breath sound measuring system was developed for the purpose to estimate the risk of CVD during sleep. Since the recorded sound signal has not only included the clear heart sounds or breath sounds but also the mixed sound with other noise, how to segment and extracting the analyzable sound signal is very important for further CVD risk analysis. The Agglomerative Hierarchical Clustering(AHC) method was proposed for classifying recording clips into four clusters to extract analyzable clips. The obtained results show that the proposed method could classify the clips into each cluster in high accuracy and is efficient to extract analyzable clips of heart sound and breath sound from a long-recorded data.

Key-Words: *Heart and breath sounds, Agglomerative Hierarchical Clustering, MFCC, Euclidean Distance*

1. Introduction

Cardiovascular disease (CVD) is a group of disorders of the heart and blood vessels including coronary heart disease, atrial fibrillation, myocardial infarction, etc. CVD is the leading cause of death globally [1]. Research shows that there is a temporal association between sleep disorders and the occurrence of CVD. People with Obstructive Sleep Apnea (usually associated with snoring) also have a higher rate to be suffered from CVD than the general public [2]. Many medical methods such as Electrocardiography (ECG) are used for monitoring the risk of CVD. However, the heart and breathing sound auscultation are one of the non-intrusive and easy ways. As smartphones or wearable devices are becoming popular due to their variety of sensors and powerful computing ability, heart and breathing sound auscultation can proceed in a home environment for CVD monitoring.

Many kinds of researches have been done to monitor CVD risk using auscultation methods. Shanti [3] presented a system that enables monitoring heart conditions by a customized stethoscope and a smartphone application. The system can be used for recording, processing, visualizing, and classification of heart sounds. Zheng [4] proposed a computer-assisted diagnosis system for intelligent diagnosis of chronic heart failure, the system uses an intelligent diagnosis model for heart sound hybrid characteristics extraction. Ohkawa [5] proposed a system to acquire and classify lung sounds into normal and pathological which takes into account the influence of heart sounds, the system achieved a higher classification rate compared with conventional methods.

These researches provided various methods for monitoring CVD, but these data are almost clear heart sound or breath sound with few other noise. A new system was proposed to monitor CVD during sleep through heart sounds and breath sounds analysis, and a clustering method was proposed to extract analyzable clips from a long-recorded data, which could be treated as

Received: 2022/06/05, Accepted: 2022/06/17
*Corresponding author: Lurui Wang
E-mail address:lurui.jp@gmail.com

the suitable signal for advanced analysis to predict the CVD risk.

2. Method

The proposed method consisted of 5 main steps. The first step is data acquisition, the heart sounds and breath sounds were recorded with a chest piece and a digital recorder. The second step is preprocessing, the audio file was filtered and segmented into clips. The third step is feature extraction, Mel-scale Frequency Cepstral Coefficients (MFCC) were extracted as a feature to represent each segmented clip. The fourth step is clustering. The Euclidean distance was calculated between each clip and formed a distance matrix, all clips were clustered into clusters with AHC meanwhile the dendrogram was built to show the structure of the audio file. The fifth step is cluster analysis, optimal cluster numbers were set based on the structure of the dendrogram, and the property of each cluster can be determined by analyzing the example nodes in each cluster. The record files can be segmented into stages based on cluster results. Different kinds of parameters can be extracted from heart sound clusters and breath sound clusters to monitor CVD risk. The schematic of the whole system was illustrated in Fig.1.

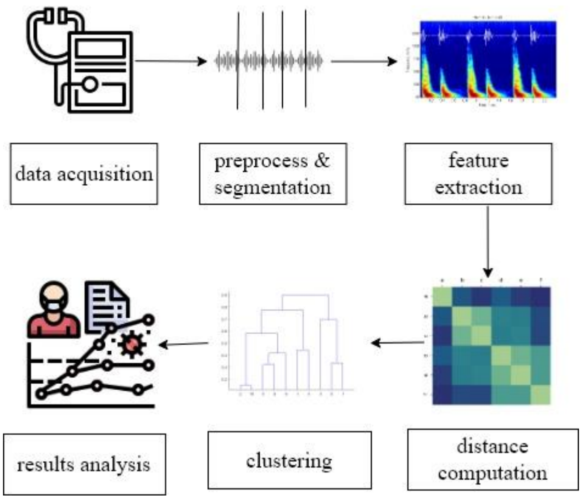

Fig.1 Schematic of system

2.1 Data Acquisition

The heart sound and breath sound were recorded by a small chest piece attached to the mitral position by adhesive tape. The chest piece made by aluminum alloy is light and small so that it does not affect the sleep quality. The chest piece was implemented with a microphone and connected to a Bluetooth transmitter. The sound data is recorded by a smartphone. The acquisition proceeded during the sleep. The record file format is MP3, the sampling rate is 44.1k Hz, and the bit depth is16-bit. As it is non-intrusive, it does not cause uncomfortable. The devices used in the acquisition was shown in Fig.2.

Fig.2 Chest piece and wireless recording system.

2.2 Preprocess

The record file was transformed into .wav format. The first step of preprocessing is filtering and denoising. As heart sound and breath sound energy concentrated in low frequency, a 20Hz-1kHz Butterworth bandpass filter was used to filter noise. The recording files were down-sampled to 2kHz. The second step of preprocessing is segmentation. The duration of clip length is settled by considering the micro and the macro aspect. One clip should be short enough to separate noise or respiration sound from heart sound, therefore the audio signal in one clip is stable. Meanwhile, the clip should be long enough to contain at least several heartbeat cycles. As the heart rate of a healthy person is usually between 60 to 100 beats per minute while they are resting, the record file was segmented into clips 10 seconds in length.

2.3 Feature Extraction

According to the research about the human hearing mechanism, the human ear has different hearing sensitivity to sound waves of different frequencies. The human ear has a higher resolution of low-frequency sounds than high-frequency sounds. The Mel scale is a direct mapping from the human auditory perceived frequency to the actual frequency of the sound. By converting the frequencies to the Mel scale, features can better match the human auditory perception [6]. There are several formulas proposed by researchers to perform this transform, the most acceptable one is as follows:

$$\mathrm{Mel}(f) = 2595 * \log_{10}\left(\frac{f}{700} + 1\right) \quad (1)$$

f is the frequency in Hertz.

MFCC were extracted from each clip file as the feature. The MFCC extraction algorithm usually includes windowing the signal into frames, applying the Fast Fourier Transform (FFT) on frames, filter spectrogram with Mel filter banks to get the Mel-spectrum, the Mel-spectrum was transformed into Mel-frequency cepstrum by taking the logarithm and then followed by applying the Discrete Cosine Transform (DCT) to get MFCC coefficients. The MFCC feature vector describes the power spectral envelope of a single frame. Fig.3 shows the waveform, Short-time Fourier transform(STFT) spectrum, the Mel-spectrum, and MFCC of a heart sound clip with a duration of 10 seconds.

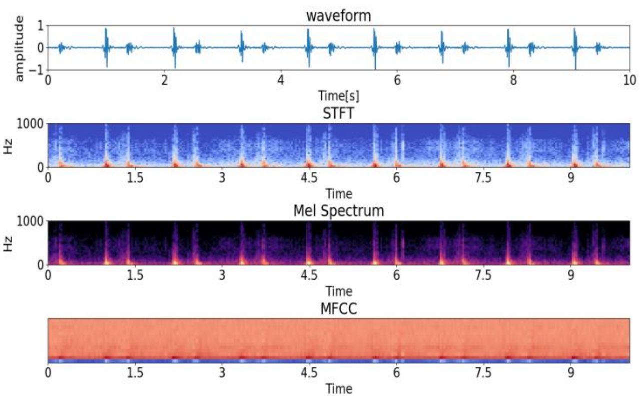

Fig.3 The waveform, STFT spectrum, Mel-spectrum, and MFCC of a heart sound clip

2.4 Similarity Calculation

The MFCC of each clip is a two-dimension matrix, each column presents for a frame, each row in the matrix corresponds to the Mel-frequency cepstral coefficients for the corresponding frame. As the heart sound signal is quasi-periodic, the MFCC matrix can be averaged by each row to get a one-dimension vector. As a vector can be presented as a point in a high dimension space by its Cartesian coordinates, the MFCC matrix can be presented as points in a high dimension space. The distance between the two clips can be measured by the distance between these two points. Based on our experiences, the Euclidean distance gave the most satisfying cluster result. The Euclidean distance between two points in Euclidean space is the length of a line segment between the two points. In general, if p and q are two points in n-dimensional Euclidean space, then the distance between them can be calculated by the following equation:

$$\mathrm{d}(\mathrm{p},\mathrm{q}) = \sqrt{(\mathrm{p}_1 - \mathrm{q}_1)^2 + (\mathrm{p}_2 - \mathrm{q}_1)^2 + \cdots + (\mathrm{p}_\mathrm{n} - \mathrm{q}_\mathrm{n})^2} \quad (2)$$

2.5 Clustering

Hierarchical clustering is a method of cluster analysis that seeks iteratively merging nodes into bigger clusters (agglomerative), or divisive clustering nodes in the inverse(divisive) to build a hierarchy of all data. It can be used to discover the structure of the data set in an unsupervised way [7]. Agglomerative Hierarchical Clustering (AHC) is the most common type of hierarchical clustering. Initially, each data point is taken as an individual cluster, then pairs of clusters are successively merged until all clusters have been merged into one big cluster that contains all objects. At each iteration, two nodes or clusters which have the minimum distance are merged, and the distance between the new cluster and other clusters needs to be updated by the linkage function. Different distance calculation criteria gave different cluster results. In the proposed method, the unweighted pair-group method with arithmetic(UPGMA) algorithm was used as the linkage function to calculate the distance between each cluster [8]. The result is a tree-based representation of all the objects, named dendrogram. The optimal number of clusters can be determined based on the structure of the dendrogram and application requirements.

3. Experiment and Result

Student volunteers were selected to participate in experiments. The experiment was proceeding during sleep. The recording files were processed with python 3.7. A 150 minutes' length file (2.5 hours) was selected from all data and segmented into 10 seconds length clips, therefore 900 clips were used in the experiments. The STFT spectrum window length is 64ms with an overlap of 32ms. The 40 Mel-scale filters were set in MFCC extraction; the distance matrix size is a symmetry matrix with a size of (900,900). The dendrogram of the clustering result was shown in Fig.4. Based on the structure of the dendrogram, the dendrogram was divided into 4 clusters. The cluster1, cluster2, cluster3, and cluster4 are presented with red, green, blue, and black respectively. The clip number of cluster1, cluster2, cluster3, and cluster are 22, 463, 3, 555 respectively.

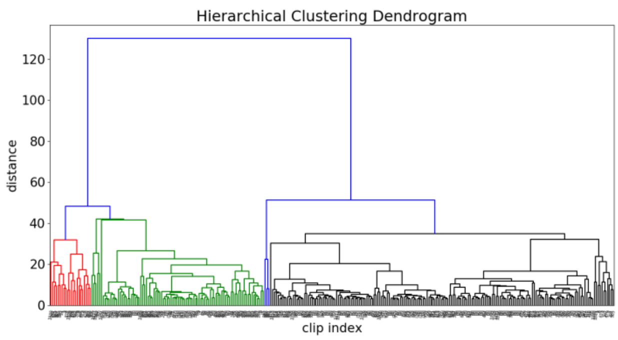

Fig.4 The dendrogram of the clustering result

As many researchers suggested that the 13-dimension MFCC coefficients used in Automatic Speech Recognition (ASR) achieved good performance [9], we use 13-dimension MFCC coefficients in the experiments. To visualize the distribution of all points in a high dimension. The Principal Component Analysis(PCA) was used for dimension reduction. PCA reduces dimensions by transforming a large set of variables into a smaller one that still contains most of the information in the large sets [10]. The MFCC vector of each clip is transformed from 13-dimension to 3-dimension using PCA therefore all MFCC vectors can be visualized in the 3-dimension space. The visualization of all feature vectors is shown in Fig.5. Each point is present for a clip. Data points were shown with different colors corresponding to the colors used in the dendrogram.

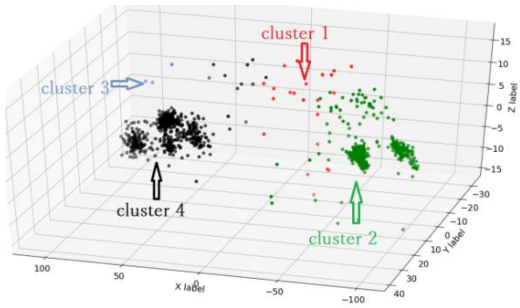

Fig.5 The 3D visualization of feature space

One clip was chosen from each cluster as an example for analysis. The waveform and Mel-spectrum of examples present for each example were shown in Fig.6. Each cluster can be analyzed by a different method based on its property.

cluster1: Clips in cluster 1 were heart sounds contaminated by noise, it can be regarded as the mixture (or transition state) of cluster 2 and cluster 3. This can also be verified in Fig.5 where the red points are scattered near the edge between cluster 2 and cluster 3. The further denoise method is needed to remove noise. This part is analyzable after denoising.

cluster 2: Most clips in cluster 2 are high-quality heart sounds, this part is analyzable and can be used for subsequent analysis directly. Cluster 1 and Cluster 2 can be used for heart health condition parameters and the estimation of CVD risk.

cluster 3: Most clips in cluster 3 were noise caused by body movement or ambient sound, as heart sound is covered by noise and hard to detect in this cluster, this part can be removed or discarded.

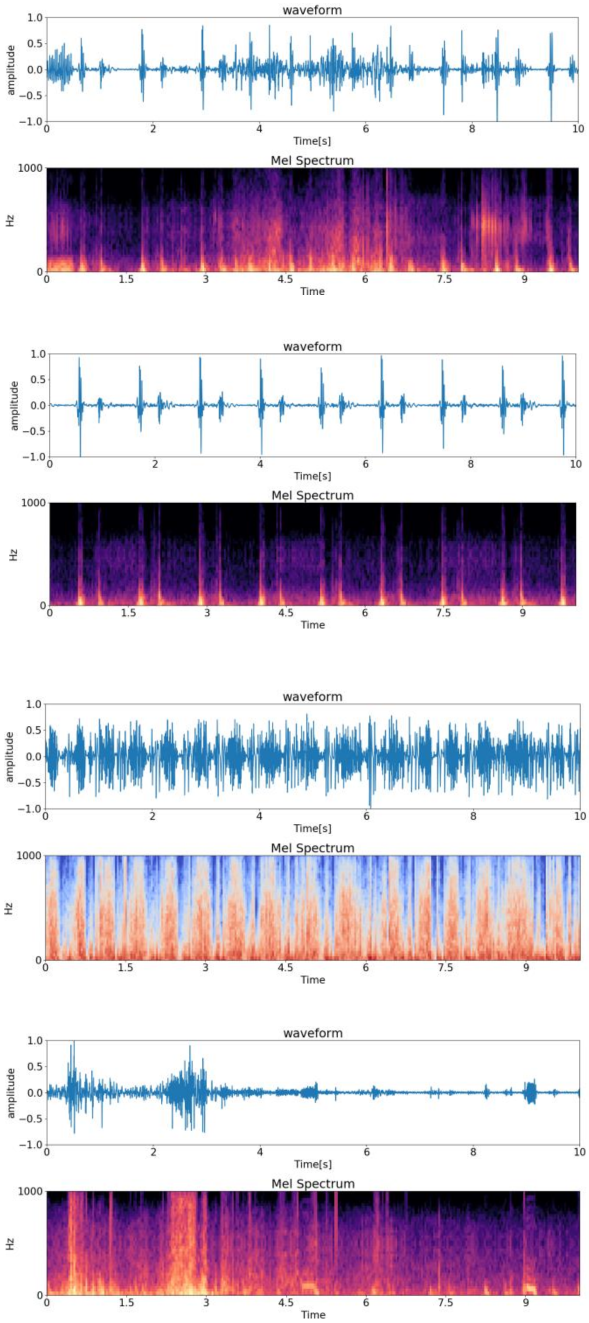

Fig.6 Typical spectrum of cluster 1, cluster 2, cluster 3 and cluster 4(from top to bottom)

cluster 4: Most clips in cluster 4 were heart sounds mixed with lung sounds or snoring. The characteristic of this part is that there are frequency components between 50-2500Hz last for more than 1 second. As Obstructive Sleep Apnea is related with CVD risk, this part can be used for the detection of apnea risk.

The waveform and Mel-spectrum of examples present for each cluster were shown in Fig.6. Each cluster can be analyzed by a different method based on its property.

For the concise layout of tables, the abbreviation used in the analysis were listed in Table 1. The characteristics of each cluster are shown in Table 2. The property column shows the property and content of clusters. The analyzable grade column indicates whether this cluster can be used for analysis, the circle symbol means that this cluster can be used for analysis, while the cross symbol meant the opposite. The analyzable content column shows which parameters can be extracted from this cluster for the prediction and monitoring of CVD risk.

Table 1 Cluster result

Abbreviation	definition
H	Heart sound
N	Ambient noise
S	Snoring sound
B	Breathing sound

Table 2 Characteristics of each cluster

Cluster NO.	property	analyzable grade	Estimation of
1	H with N	△	heart rate murmur systole diastole
2	H	○	
3	N	✖	/
4	H with N or S	○	apnea risk

The cluster distribution of all clips was shown in Fig.7. The X-axis is the clip index; each point represents for a 10-seconds-duration clip. The Y-axis is the cluster No. from 1 to 4.

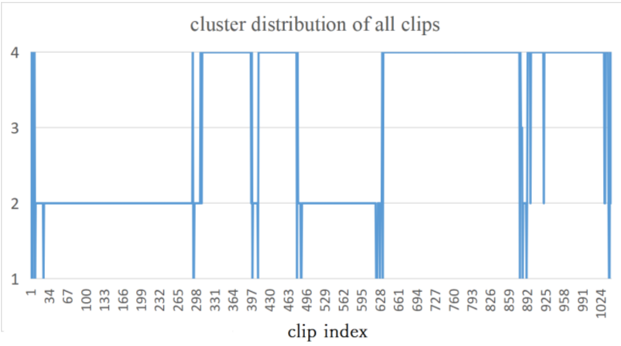

Fig.7 The label change of total record file

4. Conclusion

A new system was developed for purpose of estimation of CVD risk through heart sound and breath sound during sleep. The AHC method was proposed for extracting analyzable sound clips from a long-recorded data. The recorded data was segmented into clips each in 10 seconds length and MFCC was applied to each clip for obtaining the feature vector. AHC was performed on all clips to classify them into four clusters. The obtained results have shown that the proposed method is efficient to extract analyzable clips of heart sound and breath sound from a long-recorded data during sleep, which could be treated as the suitable signal for advanced analysis to predict the CVD risk.

References:

[1] Aggarwal S, Loomba R S, Arora R R, et al. Associations between sleep duration and prevalence of cardiovascular events. Clinical cardiology, Vol.36, No.11, 2013, pp.671-676.

[2] Cao Y, Xu Y H. Effects of sleep disorders on cardiovascular disease. Vol.25, No.1, 2020, pp.86-88

[3] Thiyagaraja S R, Dantu R, Shrestha P L, et al. A novel heart-mobile interface for detection and classification of heart sounds. Biomedical Signal Processing and Control, Vol. 45, 2018, pp.313-324.

[4] Zheng Y, Guo X, Qin J, et al. Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics. Computer methods and programs in biomedicine, Vol.122, No.3, 2018, pp.372-383.

[5] Ohkawa K, Yamashita M, Matsunaga S. Classification between abnormal and normal respiration through observation rate of heart sounds within lung sounds. 26th European Signal Processing Conference (EUSIPCO). IEEE, 2018, pp:1142-1146.

[6] Tiwari V. MFCC and its applications in speaker recognition[J]. International journal on emerging technologies, Vol.1No.1, 2010, pp.19-22.

[7] Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol.2, No.1, 2012, pp. 86-97.

[8] Le Bel F. Agglomerative Clustering for Audio Classification using Low-level Descriptors, Research Report, 2017

[9] Dave N. Feature extraction methods LPC, PLP and MFCC in speech recognition. International journal for advance research in engineering and technology, Vol.1, No.6, 2013, pp.1-4.

[10] Jolliffe I T, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol.374, No.2065, 2016