Gender Detection with Heart Sound Using MFCC-based Statistical Features

Ferda Dal, I. Yucel Ozbek

Department of Electrical and Electronics Engineering Ataturk University, Turkey
ferda_akarsu@hotmail.com, iyucelozbek@gmail.com

Abstract

This paper investigates the utility of various MFCC (Mel frequency cepstral coefficients)-based statistical features extracted from heart sound signal for the purpose of gender detection. The proposed method consists of three steps as feature extraction, support vector machine (SVM) training and testing unknown subjects. First, MFCC features are extracted from heart sounds, their various statistics are calculated and combined together to construct a new augmented features vector. Second, the statistical models based on SVM are trained by these new feature vectors. Finally, the unknown subject is tested by the proposed system to infer a decision about his or her gender. Experiments have been evaluated on a publicly available database from 98 (49 females and 49 males) subjects. The results show that the use of new combined feature vectors increased average correct classification rate from 91.83 to 93.87.

1. Introduction

Biometric recognition is a safe and effective method of recognizing people using biometrics obtained from individuals. These systems are much more advantageous than information and identity based systems such as passwords and cards. Today, many proven biometrics such as face, fingerprint, palm, vein, retina, speech have been actively used in our daily lives [1, 2, 3]. In addition to these, there are new promising biometrics, such as ECG, heart sound, and writing style, which have not yet proved themselves fully [4]. The heart sound among these new biometrics were originally used by doctors to inform people only about their health (or illnesses). However, studies conducted in recent years have revealed that heart sound also contain unique biometrics to distinguish people from each other [5]. The most important advantage of using the heart sound as a biometric is its robustness against decisiveness since it is difficult to produce a sound similar to that of particular person's heartbeat [5]. There are various studies in the literature to identify the gender of people using different biometrics. For example, the identification of the gender of a person with a facial image taken from the camera or a voice from the phone is gaining importance in different applications. Our previous research effort for gender detection of unborn babies based on their heart sounds taken on their mothers' body was to set a goal to obtain such an effective and applicable result described above. For this purpose, we first initially demonstrated how gender recognition task can be performed among adult individuals based on their heart sounds using different machine learning algorithms [6,7].

In the current study, the performance of the previously proposed system is improved with the use of new features. These new features are based on the different statistics calculated from the MFCC features. For this purpose, various statistics such as average, standard deviation, maximum and minimum values are calculated based on MFCC features. Then by combining these statistics, a new feature set of larger dimension is obtained.

The structure of this paper is as follows: In section 2 a brief information about heart sound is given. In section 3, the proposed feature extraction method and machine learning are described in detail. In section 4, the results and discussions obtained from the experimental studies are given. The main conclusions of this study are presented in section 5.

2. Heart Sounds

The main task of the heart is to pump blood into the small circulation and large circulation systems of human body. During this circulation, the blood carries the oxygen to cells and organs, and then returned back to heart with low oxygen level. To restore its oxygen level blood is pumped into the lungs, and the oxygenated blood returns back to heart. Top left and right parts of the heart where the veins are connected are called atriums while the lower parts of the heart that pump blood to our body are called ventricles. There are valves between the atria and the ventricles, and there are also valves between the ventricles and the ventricules. These valves provide one-way flow of blood. The work of the heart synchronized with the electrical stimulation of the sinoatrial knot located at the upper part of the right atrium, followed by contraction and relaxation of the atrium and ventricles. The atria and ventricles work in opposite directions, that is, one contracts and the other relaxes. This constitutes the driving force of the blood. At the time of contraction (systole), each chamber is filled with blood at the moment of relaxation (diastole) while they pump the blood out. [8]. Heart sounds are the resultant sounds of the opening and closing of the valves in the heart during heartbeat cycles. In addition, changes in the movement of the blood in the cardiovascular system are also called heart sounds [9]. Two sounds are heard in the normal heart as 'lub' and 'dub'. 'Lub' shows the first heart sound (S1) and 'dub' shows the second heart sound (S2). The S1 heart sounds is produced at the end of the stimulation of the atria and at the beginning of the stimulation of the ventricles. S2 occurs at the end of the stimulation of the ventricles. S1 is the highest peak, and its duration and frequency are between 50-100 ms and 30-100 Hz, respectively. The duration of S2 varies between 25-50 ms and the frequency varies between 100-200 Hz [10]. As an example a heart sound taken from one subject is shown in Fig.1.

3. The Proposed Method

The general block diagram of the proposed system design is shown in Fig.2. This system is composed of feature extraction, computation of statistics and classification steps which are all explained in detail as follows.

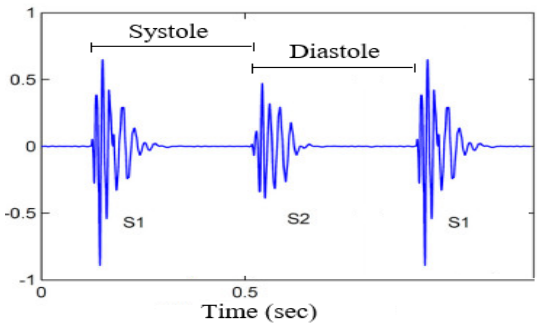

Fig. 1. Normal heart sound signal

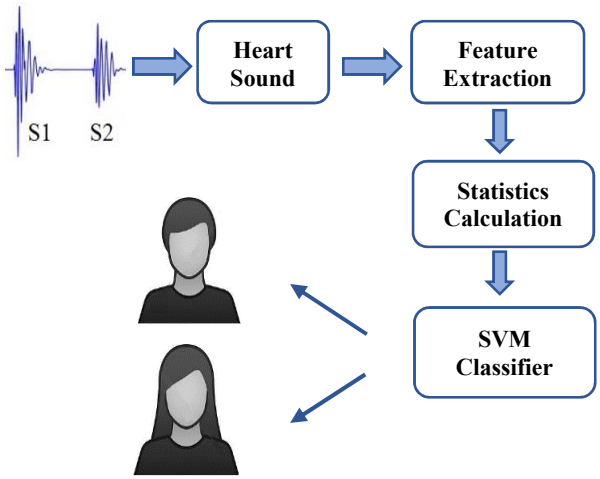

Fig. 2. General Block Diagram of Gender Detection System Design

3.1. Feature Extraction

In this study, Mel frequency cepstral coefficients (MFCC) [11] and energy features are used, widely used in the literature for speech and speaker recognition. Calculation of MFCCs is shown in Fig. 3 and each step is described as follows,

- Firstly, the heart sound signal is divided into overlapping frames.
- Then, windowing is performed to remove discontinuities at the beginning and end of the frames.
- In the following step, the DFT (Discrete Fourier Transform) of each frame is calculated.
- Then, the frequency-domain heart sound signals are passed through the mel-scaled triangle filter bank followed by a logarithm step.
- As the last step, DCT (Discrete Cosine Transform) of each frame is performed to obtain MFCC [11].

3.2. Computation of MFCC Feature Statistics

After calculation MFCC of heart sound signals, various statistics are calculated from MFCC along the temporal axis (between the frames). In this work, the eight statistics are calculated as defined in Table 2. These statistics are combined into a single augmented vector to fed a classifier for gender detection.

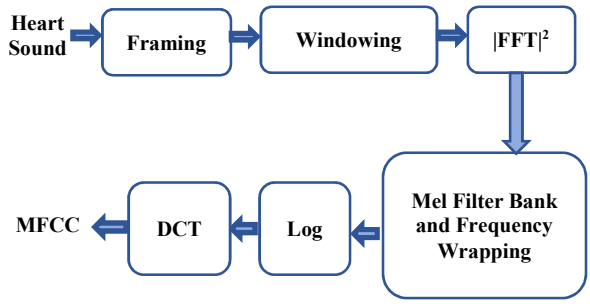

Fig. 3. Block diagram of MFCC features extraction process

Table 2. The different statistics of the MFCC

M	The average of the MFCC feature vector
Std	The standard deviation of the MFCC feature vector
Max	The maximum value of the MFCC feature vector
Min	The minimum value of the MFCC feature vector
Skew	The skewness value of the MFCC feature vector
Kurt	The kurtosis value of the MFCC feature vector
Med	The median value of the MFCC feature vector
Range	The difference between the largest and smallest value of the MFCC feature vector

3.3. Support Vector Machines (SVM) Classifier

SVMs are one of the machine learning algorithms commonly used in binary classification [12,13]. In this study we used SVM algorithm for gender classification defined as follows.

In SVM algorithm, classification is performed by defining a hyperplane which is passing through the center of the closest samples of each classes, called support vectors. The training data is defined as $S = \{x_i, y_i\}, i = 1, \dots, n; x_i \in R^d$ where i is the index number, d is the dimension, and $y_i \epsilon \{-1,1\}$ is the label of each sample. It is aimed to determine the optimal class separating hyperplane $f(x)$ defined as follows:

$$f(x) = \omega^T \Phi(x) + b \tag{1}$$

where **ω** and b denotes the normal vector to the separating hyperplane and the bias, respectively. If the transformation from the input space to the feature space is defined by $\Phi(x)$, then the optimization problem that needs to be solved to define hyperplanes is given as follows:

$$min\|\omega\|^2 + C\sum_{i=1}^{n} \xi_i \tag{2}$$
$$y_i[\omega^T \Phi(x_i) + b] \geq 1 - \xi_i \text{ and } \xi_i \geq 0,\ i = 1,2 \dots, n$$

Here, C is trade-off parameter between error and margin size, $\xi_i, i = 1, \dots, n$ slack variables defined to relax the constraints of

the separable data problem. The optimization problem of (2) aims minimizing the error cost while maximizing the hyperplane margin. In the test stage, the sign of the function $f(x)$ given in (3) is calculated by SVM algorithm to determine the class of the test data.

$$f(x) = sign\left[\sum_1^{N_s} \alpha_i y_i \Phi(s_i)^T \Phi(x) + b\right] = sign\left[\sum_1^{N_s} \alpha_i y_i K(s_i, x) + b\right] \quad (3)$$

where s_i is the support vector, N_s is the number of support vectors. Kernel K is a transformation that is used to transform the data into different Euclidean space, defined as follows:

$$K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) \quad (4)$$

In this study, Gauss kernel, $K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2}$ function is used. Here $\boldsymbol{\gamma}$ is the free parameter of Gaussian kernel function, which is related to bandwidth (variance) of the Gaussian function.

4. Experimental Studies

4.1. Database

The database used in our experimental studies is that of the database used in the literature [14]. It is obtained the result of measurements taken over many years. There are a total of 205 heart sound files of which 49 are female subjects. In this study 98 (49 female and 49 male) subjects' heart sounds were used of this database. Since there are two different heart sound files, one for training and one for testing purposes, for each subject in the database, both files are used accordingly.

4.2. Performance Measurement

To measure the performance of the proposed method, the ratio of the correctly classified subject number to the total subject number, described as follows, is used [6].

$$Acc = \left(\frac{N - N_e}{N}\right) \text{x}100 \quad (5)$$

where N_e is the number of incorrectly classified subject and N is the total number of subject.

4.3. Baseline Results (Previous work)

In our previous studies [6,7] the parameter of gender detection system is optimized in terms of MFCC order, frame length and shift length. Table 1 shows that for the best accuracy of the previous system parameters of 10th order MFCC, 155 ms frame length and 100 ms shift length is used. Hence, corresponding percent accuracy rate of 91.83 is considered as the baseline of the current study.

Table 1. The baseline performance and related parameters for gender detection

Feature Type	Frame Length	Shift Length	Acc (%)
MFCC-10 + Energy	155	100	91,83

4.4. Experimental Results of Current Work

Several MFCC statistics were used together to examine the performance of the system. These statistics were concatenated to form a single augmented vector to be used as a new feature vector and these are used in Support Vector Machine algorithm.

Table 3 shows the results obtained in this manner. From the examination of this table, it is clear that the best accuracy rate of 93,87 is obtained when the augmented vector is composed of mean, std, max, skewness, kurtosis, range vectors of the MFCC.

Table 3. Performance of the current system with SVM according to different MFCC statistics

Combination of Feature Statistic	Feature Dim.	Acc (%)
M + Std + Skew + Kurt	11x4=44	87,71
M + Std + Min + Skew+ Kurt + Med	11x6=66	90,77
M + Kurt + Med+ Range	11x4=44	89,75
M + Std + Min + Skew	11x4=44	91,79
M + Std + Max+ Skew + Kurt + Range	11x6=66	93,87

5. Conclusion

In this study, the gender of the subjects was determined by using the heart sound information. In the gender detection, the feature extraction from heart sound was performed first. Using these features; SVMs are trained for each gender class using available training data, and then trained models are evaluated using test data. According to experimental results, we obtained following main results. When the augmented feature vector that obtained from different MFCC statistic was used in SVM training and testing stages, it is observed that the accuracy rate was increased from baseline value of 91.83 to 93.87.

6. References

[1] Jain A.K., Ross A., Prabhakar S., "An introduction to biometric recognition", Circuits and Systems for Video Technology, IEEE Transactions on , vol.14, no.1, p 4-20, Jan. 2004.

[2] Wildes R. P., "Iris recognition: An emerging biometric technology", in Proceedings of the IEEE, vol. 85, no. 9, p 1348–1363, 1997.

[3] Shu W. and Zhang D., "Automated personal identification by palmprint", Optical Engineering, vol. 38, no. 8, p 2359-2362, 1998.

[4] Gafurov D., "Emerging Biometric Modalities: Challenges and Opportunities", SecTech/DRBC, p 29-38, 2010.

[5] Koksoon Phua, Jianfeng Chen, Tran Huy Dat, Louis Shue, "Heart sound as a biometric", Pattern Recognition, Volume 41, Issue 3, p 906-919, March 2008.

[6] F. Dal, I. Ozbek, ve S. Coşgun, "Kalp Sesi ile Cinsiyet Tanıma", Signal Processing and Communications Applications Conference (SIU), İnönü Üniversitesi, May, 2015.

[7] F. Dal and I. Ozbek, "Gender Detection with Heart Sound Based on SVM", Signal Processing and Communications Applications Coference (BİYOMUT), İzmir Kâtip Çelebi Üniversitesi, May, 2015.

[8] K. Phua, J. Chen, T. H. Dat, and L. Shue, "Heart sound as a biometric" , Pattern Recognition, Volume 41, Issue 3, p 906-919, March 2008.

[9] Ö. Say, " Kalp Seslerinin Analizi ve Yapay Sinir Ağları ile Sınıflandırılması", Y.Lisans Tezi, Fen Bilimleri Enstitüsü, İstanbul Teknik Üniversitesi, İstanbul, 2002.

[10] S. Debbal, "Model of Differentiation between Normal and Abnormal Heart Sounds in Using the Discrete Wavelet Transform", Journal of Medical and Bioengineering, 3 (1), 5-11, 2014.

[11] Davis S. And Mermelstein P., "Comparison of parametricrepresentations for monosyllabic word recognition in continuously spoken sentences", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 28, no. 4, p 357–366, Aug 1980.

[12] V. Vapnik, The Nature of Statistical Learning Theory, Springer,1995.

[13] E. Yakut, B. Elmas, ve S. Yavuz, "Yapay Sinir Ağları Ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini", Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 19(1), 139-157,2014.

[14] A. Spadaccini, F. Beritelli, "Performance evaluation of heart sounds biometric systems on an open dataset", Digital Signal Processing (DSP), 2013 18th International Conference on, p 1-5, 1-3, July 2013.