Human identification using heart sound

Koksoon Phua, Tran Huy Dat,Jianfeng Chen and Louis Shue
Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
Email: {ksphua,hdtran,jfchen,lshue}@i2r.a-star.edu.sg.

***Abstract*— In this paper, we investigate the possibility of using heart sound as a biometric for human identification. The most significant contribution of using heart sound as a biometric is that it cannot be easily simulated or replicated as compared to other conventional biometrics. Our approach consists of a robust feature extraction scheme which is based on cepstral analysis with a specified configuration, combined with Gaussian mixture modeling. Various experiments have been conducted to determine the relationship between various parameters in our proposed scheme. The results suggest that parameter values appropriate for heart sounds should be significantly different for equivalent parameters used in conventional cepstral analysis for speech processing. In particular, heart sounds should be processed within segments of 0.5 second and using the full resolution in frequency domain. Secondly, higher order cepstral coefficients, carrying information on the excitation, are also useful. Preliminary results indicate that with well-chosen parameters, an identification rate of up to 96% is achievable for a database consisting of 7 individuals, with heart sounds collected over a period of 2 months.**

I. Introduction

Since the last decade, reliable human authentication and identification systems have been used in many applications, such as personnel security, military, finance, airport, hospital, digital rights management systems, etc. [1]. Conventional biometric systems using behavioral and/or physiological characteristics to allow recognition of an individual, e.g. fingerprint, iris, face and voice, are becoming more popular [1]–[4]. However, a common weakness of these system is their vulnerability to the possibility to falsify these features [2,4,5].

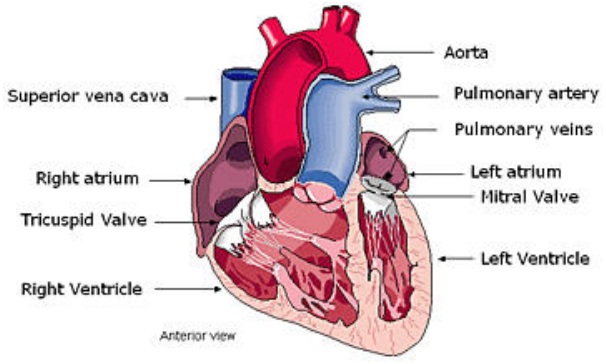

Fig. 1. Cross-section of a typical human heart

Correspondence author: Koksoon Phua. Email: ksphua@i2r.a-star.edu.sg. Tel: +65-6874 8401. Fax: +65-6776 8109

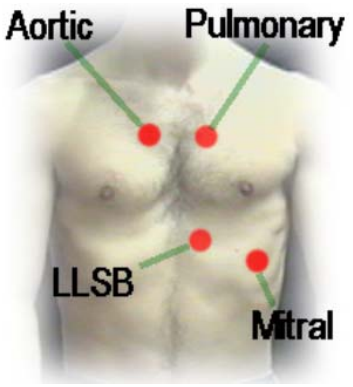

Fig. 2. Typical auscultation sites to place microphones

Studies into new paradigms in multi-modal biometrics is currently a very active research topic. New biometrics using features such as hand vascular pattern, vein, gait, human tissue, knuckle, ear canal and even evoked brain signals have been proposed [6,7]. Another new and prospective candidate for biometric is the electrocardiogram (ECG) [8,9] which yields a relative high result for human identification tasks [8,9]. However, we note that ECG for identification is generally cumbersome due to the many (at least three) electrodes required [9].

In this work, we investigate the possibility of using human heart sounds – an acoustic signal – as a reliable biometric for human identification. Human heart sounds are very natural signals, which have been applied in the doctor's auscultation for health monitoring and diagnosis for thousands of years. In the past, study of heart sounds focus mainly on the heart rate variability [10]. However, we conjecture that since the heart sounds also contain information about an individual's physiology, such signals have the potential to provide a unique identity for each person. Like ECG, these signals are difficult to disguise and therefore reduces falsification. Moreover, heart sounds are relatively easy to obtain, by placing a conventional stethoscope on the chest, for example.

II. Mechanisms for heart sound production

In this section, we will briefly outline the mechanisms involved in heart sound production.

The human heart has four chambers, two upper chambers called the *atria* and two lower chambers called *ventricles*, as shown in Figure 1. There are valves located between the atria and ventricles, and between the ventricles and the major

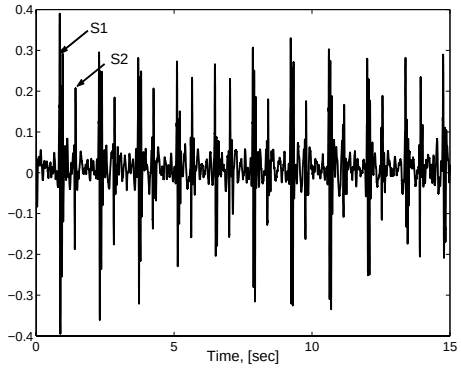

Fig. 3. Waveform of first heart sound S1 and second heart sound S2

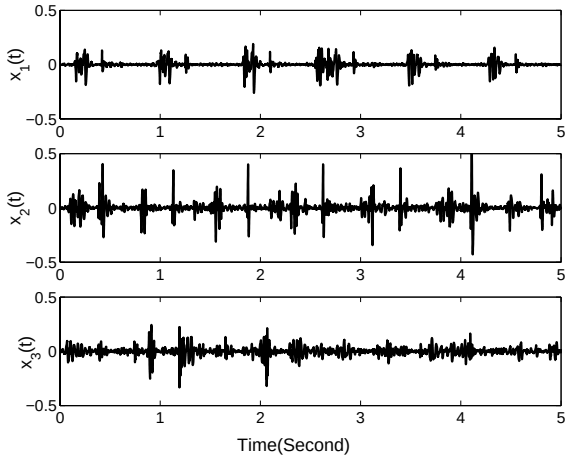

Fig. 4. Heart sound waveform comparison

arteries from the heart [11]. These valves close and open periodically to permit blood flow in only one direction. Two sounds are normally produced as blood flows through the heart valves during each cardiac cycle. The first heart sound **S1** is a low, slightly prolonged "lub", caused by vibrations set up by the sudden closure of the mitral and tricuspid valves as the ventricles contract and pump blood into the aorta and pulmonary artery at the start of the ventricular systole. The second sound **S2** is a shorter, high-pitched "dup", caused when the ventricles stop ejecting, relax and allow the aortic and pulmonary valves to close just after the end of the ventricular systole. They are the "lubb-dupp" sounds that are thought of as the heartbeat. S1 has duration of about 0.15 second and a frequency of 25-45 Hz. On the other hand, S2 lasts about 0.12 second, with a frequency of 50 Hz. Figure 3 shows the waveform of the S1 and S2 heart sound.

According to [12], the sounds associated with the opening and closing of different values can be heard by placing a microphone directly on the various auscultation points on the chest wall as shown in Figure 2. The time-series and spectral plots of regular heart sounds recorded from five different people are shown in Figures 4 and 5. While signal envelopes are similar, the details of the time- and frequency-domain waveforms are relatively different. From Figure 4, it can be seen that people may have regular or irregular heart sounds or a combination of both. Furthermore, the person with the

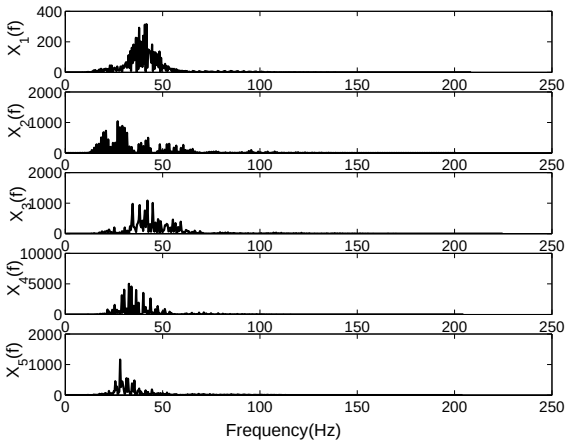

Fig. 5. Heart sound spectrum comparison

irregular heart sound exhibit more distinctive features in the waveforms. These preliminary observations further suggest a distinctiveness of using heart sound for identification.

III. Human heart sound as a biometric

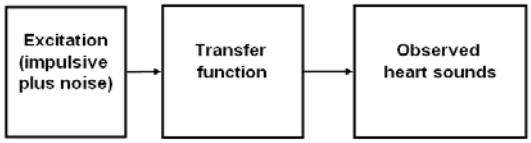

Fig. 6. Linear model of heart sound

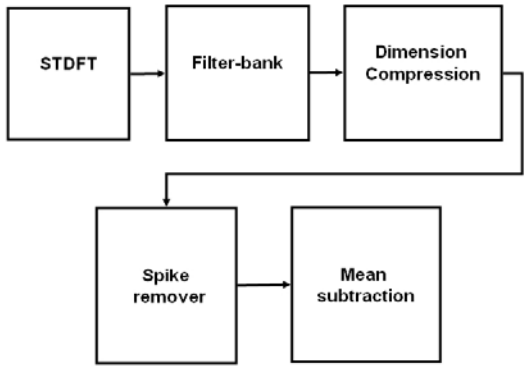

Fig. 7. Block diagram of feature extraction process

A. Choice of features

Due to the overlap of the heart sound components and the noises and artifacts caused by other internal organs, a full analysis of the heart sound in the time-domain is difficult. For the biometric application, the physiological properties of the heart sound is more important than the heart rate and accordingly we will subsequently restrict the processing of heart sound to the frequency domain. In this work, we will adopt the cepstral features, which have been used in many speech processing application, in speech recognition and speaker identification in particular. Similar to the modeling of

speech signal, we will regard heart sound as the outcome of a time-varying linear system (see Figure 6), where the excitation input carries information on the heart beat rate, and the transfer function is time varying.

We will presently describe each of the six functional blocks in the feature extraction module for heart sounds as shown in Figure 7.

1) Short-time Fourier transform: As in the case of speech signals, we will assume a short-time stationary property for heart sounds, which conveniently allows a STDFT analysis

$$X[n,k] = \sum_{m=0}^{N-1} x * w[m + (n-1)S] \exp\left(-j\frac{2\pi}{N}km\right), \quad (1)$$

where n is the frame index, k is the frequency index, N is the frame length, S is the frame shift and w denotes the window. Unlike speech signals, where the vocal track is changing after each 20-25ms, heart sounds are more stationary and therefore the window length should be larger. The optimal window length was found to be about 500ms through experiments conducted in subsequent section.

A related issue is the effect of the window shift on the performance of the pattern recognition method. When Gaussian mixture modeling is used for classification, independence between samples are required in order to accurately estimate the probability distributions. Consequently, the non-overlap windowing is optimal. On the other hand, having a small number of samples can lead to deficiencies in the distribution fitting, and in this case, e.g. when dealing with small number or short training samples, overlap should be used. We will discuss this issue in detail in Section IV.

Next, we use only the information in the spectral magnitude and ignore the phase components which are typically more sensitive to the noise.

2) Filter-bank: In this block, the signal spectrum will pass through a filter-bank set. The usage of these filter-banks are motivated by the fact that, the sound spectrum has some special shapes and are distributed by a non-linear scale in frequency domain. Using the filter-banks with spectral characteristics which are well-matched to those of the desired signal, the contribution of noise components in the frequency domain can be reduced. Mel-frequency filter-banks have been shown to be the best in the speech recognition and speaker identification. However, unlike the speech signal, heart sound spectrum is concentrated within the range of 20Hz-150Hz. In such a narrow band, full resolution is required to capture more information of the signal spectrum. In this work, heart sound spectrum is simply processed by filtering out the frequency bins outside the range of 20Hz-150Hz.

3) Dimension compression: This block is very similar to the standard MFCC feature extraction. The output spectral magnitude is compressed in the logarithm domain, followed by the discrete cosine transform which is considered to be an empirical principle component analysis. The cepstral component $c[n,k]$ can be written as

$$c[n,k] = \sum_{m=0}^{\bar{K}-1} \log(|X[n,m]|) \cos\left(\frac{km\pi}{\bar{K}}\right), k = (1,2,...,\bar{K}) \quad (2)$$

where $\bar{K}$ is the number of bins in the frequency band of 20Hz-150Hz.

The dimension reduction is achieved by selecting the first 24 coefficients for each frame. Note that the higher coefficients are considered to be contributions from the excitation process which is less informative. To distinguish heart-sound's feature from the standard MFCC, we will call this feature set the Linear Frequency Bands Cepstra (LFBC).

4) Spike removal: The quality of the heart sound recording often suffers from bursty interferences caused by the movement of the stethoscope. The spectra of these impulsive noises and heart sounds overlap and therefore conventional filtering technique are ineffective. In this work, we set an energy threshold to remove the high energy segments that contains the burst

$$10\lg E[n] - \min_n(10\lg E[n]) \geq \mu, \quad (3)$$

where n is the segment index and $\mu = 6dB$ is an energy threshold.

5) Cepstral means subtraction: Since the positions of stethoscope cannot be fixed at all time, there is always a fluctuation on the "relative transfer function" as characterized by the propagation of heart sounds to the recorder. To remove this effect, we will apply the cepstral mean subtraction. The multiplication of the signal measured in a fixed position, $X[n,k]$, and the relative transfer function in the frequency domain, $H[k]$, is equivalent to a superposition in the logarithm-domain and is given by

$$\log(|Y[n,k]|) = \log(|X[n,k]|) + \log(|H[k]|). \quad (4)$$

Consequently, the cepstra of the recorded signals can be denoted as follows

$$c_Y[n,k] = c_X[n,k] + c_H[k]. \quad (5)$$

The last component in Equation (5) can be removed by taking the long term averaging in each dimension k

$$c_{Y,k}[n] - \langle c_{Y,k}[n]\rangle = c_{X,k}[n] - \langle c_{X,k}[n]\rangle. \quad (6)$$

B. Classification scheme

The Vector-Quantization (VQ) and Gaussian Mixture Modeling (GMM) are conventional and successful methods for the speaker recognition approaches. In this work, we adopted these methods for the proposed heart-sound based identification.

The basic idea of using vector quantization is to compress a large number of cepstral vectors into a small set of code vectors. The VQ codebook is usually trained with the LBG algorithm [13] to minimize the quantization error when replacing all feature vectors with their corresponding nearest code vectors. The Euclidean distance is often used as a quantization error measure. The LBQ-VQ can be considered as an approximation of the density function by discrete points and the main advantage of this method is its ability to archive relative high identification rate with very short test signals. Other VQ methods have also been proposed by Kohonen [14] to globally optimize the codebook after they are generated with a certain unsupervised learning algorithm. This method is called learning vector quantization (LVQ). Unlike LBG-VQ,

the LVQ algorithms tends to define directly the classification borders between classes by nearest-neighbor rule. However, the identification decision in testing is based on a vector sequence rather than one individual vector, therefore, the higher correct classification rate for feature vector, achieved with the LVQ is not always improve the identification rate. moreover, the computational cost of LVQ is higher than LBQ. In this work, we have implemented the fast version of LBQ with the code number $N = 2^k$.

The second pattern matching method implemented in this work is the Gaussian mixture modeling (GMM) [15], which can be considered as a stochastic generalization of the vector quantization. In contrast to LBQ-VQ, the GMM method approximates the continue probability density function by a mixture of Gaussian pdfs

$$p(\vec{x}|\lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}) \tag{7}$$

where $\vec{x}$ is a D-dimensional random vector, $b_i(\vec{x})$, i = 1,...,M, are the component densities and p_i, i = 1,...,M, are the mixture weights. Each component density is a D-variate Gaussian function of the form

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2}\,\Sigma_i^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)\right\} \tag{8}$$

with mean vector $\vec{\mu}_i$ and covariance matrix Σ_i. The mixture weight satisfies the constraint that $\sum_{i=1}^{M} p_i = 1$ The complete Gaussian mixture density is parameterized by the mean vector, covariance matrix and mixture weights from all component densities. These parameters are collectively represented by the notation

$$\lambda = \{\, p_i, \vec{\mu}_i, \Sigma_i\} \;\; i \;=\; 1, \ldots, M. \tag{9}$$

The expectation maximization (EM) algorithm is usually due to its simplicity and quick convergence. Note that, the LBQ-VQ codebook is use in this work as an initial for the GMM fitting by EM algorithm. For identification using heart sound, the GMM model λ is trained for each person and then in testing, each heart signal is referred to by his/her model as the maximum of the likelihood (conditional probability) measures.

IV. Experimental results and discussions

In Section IV-A, we will provide a detailed description of the data collection and experimental procedure. The evaluation of our proposed biometric system using different parameters will then be discussed in Section IV-B. Finally, in Section IV-C, we will briefly comment on the performance when fewer training and testing data samples are available.

A. Experimental setup

The heart sound signal was recorded using a Welch Allyn Meditron Electronic Stethoscope, as shown below in Figure 8. The electronic stethoscope was placed on the chest of the participant seated in a relaxed position. For this experiment, seven sets (5 males, 2 females) of data was collected: 100 heart sound readings recorded by each participant over a two-months

Fig. 8. Welch Allyn Meditron Electronic Stethoscope

period. The interval between each recorded heart sound was required to be at least one hour apart, and the stethoscope was required to be placed at the same location on the chest for all readings.

B. Identification performance

During the training phase, 6 heart sounds from each person recorded on different days were used to build the models. During testing, 70 heart sounds were randomly chosen from each person. Two methods of Vector Quantization (VQ) and Gaussian Mixture Modeling (GMM), described in Section III, were implemented. The linear frequency band cepstra (LBFC) was used as the main feature for training and testing.

In the following, we will evaluate the performance by varying in turn 1) number of features, 2) number of mixture components, 3) comparing LBFC against MFCC, 4) frame length and frame shift.

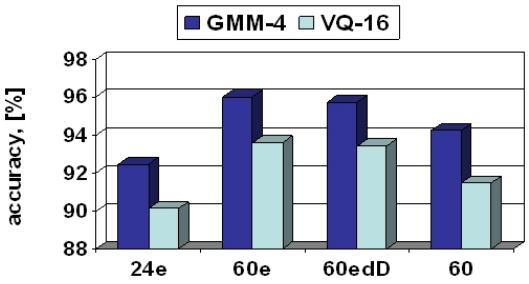

Fig. 9. Evaluation results of Gaussian Mixture Modeling using 4 mixture (GMM-4) and Vector Quantization using 16 vector dimension (VQ-16) at four different configurations, namely (a) 24 first cepstra plus log-energy (24e), (b) 60 cepstra plus log-energy (60e), (c) 60 cepstra plus log-energy, delta and acceleration coefficients (60edD), and (d) 60 cepstra without log-energy (60)

1) Number of features: The four configurations using different combinations of features tested are

1) 24 first cepstra plus log-energy,
2) 60 cepstra plus log-energy,
3) 60 cepstra plus log-energy, delta and acceleration coefficients, and
4) 60 cepstra without log-energy.

As shown in Figure 9, the GMM-based system achieved the best result of 96.01% using the configuration with 60 cepstra plus log-energy configuration. For the VQ-based system, the best score was 93.58%, also using the same set of features.

Unlike in speech recognition, where the higher order cepstra which carry information regarding the excitation component

and are generally not useful, these components were observed to be useful for the heart sound analysis. In both VQ-based and GMM-based systems, the 60-cepstra systems give approximately 3% higher accuracy than the 24-cepstra ones. In addition, the use of additional delta and acceleration coefficients do not improve the performance of the system. This is very likely due to the fact that the vocal track has a bigger dynamic range than the equivalent for human heart. The log-energy was shown to be important and made a contribution about 2% in the accuracy rate. Note that, here the frame length of 512ms without overlap has been used. We will discuss the choice of these parameters later.

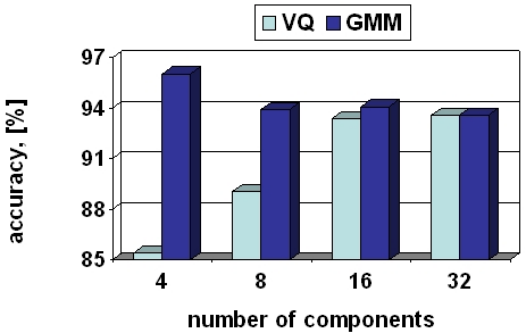

Fig. 10. The performance of the VQ and GMM methods based on the number of mixutre components

2) Number of mixture components: In this section, we will determine the optimal choice of the number of mixture components. As shown in Figure 10, the performance of GMM-based system is the best when using 4 components and actually degrade when the number of components increases further. In contrast, the VQ-based system requires more components and reach its best at 16 components. This is likely because the discrete approximation of the probability density function by the VQ requires more data points than using the continue Gaussian components. Based upon this consideration, we conclude that the GMM-system is better in both the accuracy rate and computational cost.

3) Comparison to MFCC: One additional interesting question we were interested in was: what is the impact on overall performance if the standard MFCC features were used instead? In Section III-A, we argued that the LBFC is more suited to the current task which will be backed up by some experimental evaluations. As shown in Figure 11, the LBFC results overcome the MFCC's by more than 3% in both the VQ-based and GMM-based systems. Note that LBFC and MFCC are calculated using the same frame length and shift.

4) Frame length and shift: As mentioned in Section II, the human heart sound is quasi-stationary with much longer segment than speech and consequently the standard 20-25ms frame length for speech processing is not well-matched. This point was born out by our experiments, as shown in Figure 12. From this figure, the frame length is shown to be very important issue. The standard in speech frame length is ineffective for the heart sounds, yielding less than 70% accuracy. The 512ms is shown to be the optimal choice of frame length. Next, we evaluate the systems with 512 ms frame length but different frame shifts as shown in Figure 13. The results show that, the effect of the frame shift is less than frame length and the non-overlap system yields the best results, particularly in the GMM-based system. This result also conform the consideration of that, for the distribution fitting system, the overlap manner is unnecessary since it produces a dependence between samples. Note that, the non-overlap property is not suitable when test sequences are too short, as this can result in very few number of samples.

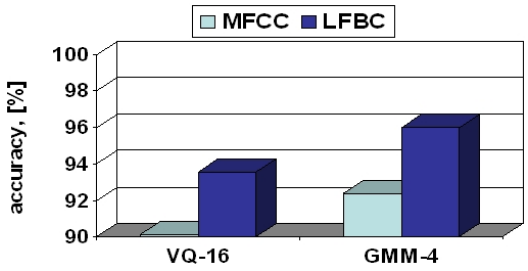

Fig. 11. The performance of the VQ and GMM methods when using the (a) LFBC feature and (b) MFCC feature

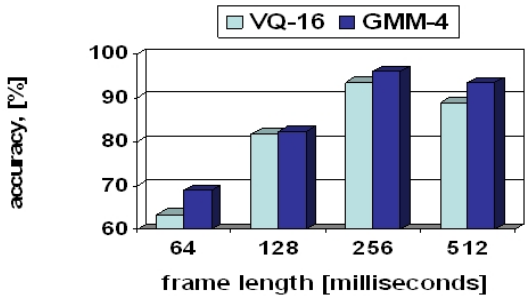

Fig. 12. The performance of the VQ and GMM methods when using different frame length

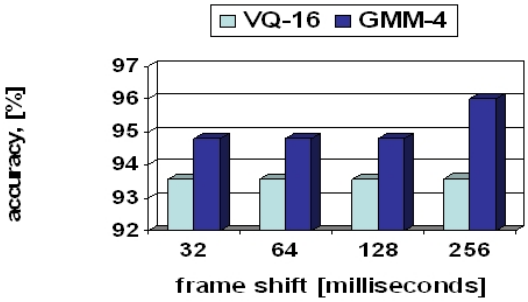

Fig. 13. The performance of the VQ and GMM methods when using different frame shift at a frame length of 256ms. Note that when frame shift = 256ms, there is no overlap.

C. Evaluation with shorter training/testing samples

In this section, we will address the question of the identification performance with respect to the length of the training data and the testing data. The models were trained using data of 10, 30, 60, 210 and 420 seconds in duration; the trained models were then tested using samples of 1, 3, 5, 10, 30 and

TABLE I

EVALUATION WITH VARIOUS LENGTHS OF TRAINING AND TESTING DATABASES USING THE GMM METHOD

Training length	Test length				
	1sec	5sec	10sec	30sec	60sec
10sec	68.75%	76.22%	79.69%	78.65%	82.64%
30sec	76.39%	89.58%	91.49%	92.71%	92.49%
60sec	78.89%	89.41%	91.67%	92.71%	93.06%
210sec	81.60%	90.97%	92.71%	94.27%	94.79%
420sec	80.38%	90.80%	93.92%	95.66%	96.01%

60 seconds of heart sounds in the same manner as described earlier.

The identification results for each training and testing data length using GMM classification method are shown in Table I. From the table, we observed that the identification performance is higher than 90 % when the training and testing data lengths are more than 30s and 5s respectively. The optimal testing length seems to be about 30s and only very small difference between this and 60s length test.

V. CONCLUSIONS

In this paper, we have investigated the possibility of using heart sound in the human identification task. The most salient feature in using the heart sound as a biometric is that it cannot be easily simulated or copied, as compared to other biometrics such as face, fingerprint or voice. Furthermore, the proposed framework is relatively economical to install and maintain as it requires only an electronic stethoscope and a simple processor and database server for carrying out the identification task. The proposed cepstral features and the associated pre-processing have been shown to be suitable for the heart sounds in the human identification task, yielding a relatively good and robust results in the experimental evaluation.

Work is currently under way to answer two additional questions: 1) the scalability of using heart sound as a biometric for a larger database? 2) the invariance of the proposed features over significant time intervals.

REFERENCES

[1] J. Ortega-Garcia, Bigun J., D. Reynolds and J. Gonzalez-Rodriguez. "Authentication gets personal with biometrics", *Signal Processing Magazine, IEEE* , vol. 21 , issue 2 , pp. 50 - 62, March 2004.

[2] A. K. Jain, A. Ross, and S. Prabhakar. "An Introduction to Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4 – 20, January 2004.

[3] J. P. Campbell Jr. "Speaker Recognition: A Tutorial", *Proceeding of the IEEE*, vol. 85, no. 9, pp 1437-62, September 1997.

[4] L. O. Gorman. "Comparing Passwords, Tokens, and Biometrics for User Authentication," *Proceedings IEEE*, vol. 91, no. 12, pp. 2019-2040, Dec. 2003.

[5] M. Faundez-Zanuy. "On the vulnerability of biometric security systems", *Aerospace and Electronic Systems Magazine, IEEE*, vol 19, issue 6, pp. 3- 8, June 2004.

[6] D. Peter. "Biometrics continue to evolve", *Biometric Technology Today*, vol. 11, issue 9, pp 7-8, September 2003.

[7] R. Palaniappan and P. Raveendran. "Individual identification technique using visual evoked potential signals", *IEE Electronics Letters*, vol.138, issue 25, pp.1634-1635, December 2002.

[8] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, B. K. Wiederhold. "ECG to identify individuals", *Pattern Recognition*, vol. 38, no. 1 , pp. 133-142, January 2005.

[9] L. Biel, O. Pettersson, L. Philipson, and P. Wide. "ECG Analysis: A New Approach in Human Identification", *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808 – 812, June 2001.

[10] M. Malik. "Heart rate variability: standards of measurement, physiological interpretation, and clinical use", *European Heart Journal* Circulation 93 (**??**), pp. 1043–1065, 1996.

[11] F. G. William. *Review of Medical Physiology*, Prentice Hall, 1997.

[12] B. N. Robert. *Noninvasive Instrumentation and Measurement in Medical Diagnosis*, CRC Press, 2002.

[13] Y. Lindo, A. Buzo, and R. M. Gray. "An algorithm for vector quantizer design," IEEE. Trans. Communication, vol. COM-28, pp. 84-95, Jan. 1980

[14] T. Kohonen. "Self-Organizing Maps", *Berlin: Springer-Verlag*,1997.

[15] D. A. Reynolds and R. C. Rose. "Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models.", *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 72 – 83, January 1995.