Performance Evaluation of Heart Sounds Biometric Systems on an Open Dataset

Andrea Spadaccini, *Graduate Student Member, IEEE* and Francesco Beritelli, *Member, IEEE*
Dipartimento di Ingegneria Elettrica, Elettronica ed Informatica
University of Catania, Italy
{andrea.spadaccini,francesco.beritelli}@dieei.unict.it

Abstract

Recently, many systems and approaches that employ heart sounds as physiological traits for biometric recognition have been investigated. However, those systems are often tested on small, diverse and closed datasets, making it difficult to compare their performance. In this paper, we present HSCT-11, an open dataset containing data collected from 206 people that can be used for performance evaluation of heart sounds biometric systems, and we use it to benchmark two such systems. The most performing one shows an Equal Error Rate of 13.66 % on this database, a result that will be the baseline for all the future evaluations made using this dataset.

1. Introduction

In the last 4 years, many researchers have investigated the possibility of using heart sounds as physiological traits for biometric recognition [1, 2, 6, 8, 9, 11, 14, 15].

However, as it is common for new fields of research, those studies often report their results using different performance metrics, and the heart sound databases are often heterogeneous and unavailable to other research groups.

In this paper, we present HSCT-11, an open heart sounds database that is freely available to the research community. To the best of our knowledge, it is the largest database in terms of the number of people. We then proceed with the description of two biometric systems based on heart sounds and we use the HSCT-11 database to compare their performance, expressed in terms of Equal Error Rate.

The paper is structured as follows: in Section 2 we introduce the concept of heart-sounds biometry; in Section 3 we describe the related works; in Section 4 we describe in more detail two biometric systems based on heart sounds; in Section 5 we introduce the HSCT-11 database and we analyze how the two biometric systems perform on this database; finally, in Section 6 we draw our conclusions and describe possible future work on the topic.

2. Heart-sounds Biometry

2.1. Physiology of Heart Sounds

The heart sound signal is a complex, non-stationary and quasi-periodic signal that is the audible result of the continuous pumping work of the heart. It is composed of several beats, or cycles, that contain different smaller sounds which are produced by the movement of the heart valves and by the blood flowing.

The primary sounds are called S1 and S2. They are produced, respectively, by the closure of the tricuspid and mitral valves and by the closure of the aortic and pulmonary valves. The other sounds are S3 and S4, usually quieter, and murmurs, high-frequency noises that are usually pathological.

We decided to use in our systems only the two primary sounds, because they are more universal and way louder than other sounds. We separate them from the rest of the signal before the feature extraction phase, using the algorithm described in Section 4.1.

2.2. Comparison with other Biometric Traits

Jain et al. [10] present an evaluation of the most widespread biometric traits according to 7 qualities. In this section, we will evaluate heart sounds with respect to the same qualities.

- **Universality:** *high*
  Every living human being must possess a functioning heart, so the heart sound signal can be acquired from anyone.
- **Distinctiveness:** *medium*
  While encouraging, the actual systems' performance is still far from the most established traits, and even if some results are similar to other traits (like voice), the database used for performance evaluations is still not large enough.
- **Permanence**: *low*
  So far, there are no studies that analyze how the biometric properties of heart sounds change over time.

- **Collectability**: *low*
  The novelty of the trait also implies that - so far - there are no specialized sensors for the acquisition of heart sounds; this means that normal electronic stethoscopes used in medical environments must be used for the acquisition phase. The collection process is thus not easy, it requires a careful positioning of the stethoscope and the direct contact with the chest skin of the person that must be authenticated.

- **Performance**: *low*
  The accuracy of the systems, as discussed before, should be improved; also, the algorithms are still computationally expensive and would need to be optimized for real-world usage.

- **Acceptability**: *medium*
  Heart sounds are probably perceived as unique and trustable, but people might be unwilling to use them in daily authentication tasks;

- **Circumvention**: *low*
  It is very difficult to reproduce the heart sound of another person, and it is also difficult to record it covertly in order to reproduce it later.

The main advantages of heart sounds are, so far, the High Universality and the Low Circumvention.

The first point is undeniable and objectively true. This property is shared with all the biometric traits that depend on organs whose functioning is crucial for our life, like the brain. This also means that heart sounds - needing cooperation from the subject - cannot be used in situations like crime scene analysis.

The second point derives from the difficulty of recording heart sounds. A malicious user of the system would have trouble in recording another user's heart sounds for using them later, not mentioning the difficulties in hiding a player that should be used during the authentication phase. It must be noted that, since most systems use generative models as identity templates, if the models are not properly secured the same templates could be used to generate fake heart sounds that would match very well the template itself.

The main drawbacks of heart-sounds biometry are probably the Low Performance and, above all, its overall immaturity as a biometric trait. Of course, heart-sounds biometry is a new technique, and as such many of its current drawbacks will probably be addressed and resolved in future research work.

3. Related Work

In Table 1 we summarized the main characteristics of some of the most relevant papers written about heart-sounds biometry, using the following criteria:

- **Database** - the number of people involved in the study and the amount of heart sounds recorded from each of them;

- **Features** - which features were extracted from the signal, at frame level or from the whole sequence;

- **Classification** - how features were used to make a decision.

Ref.	Database	Features	Classification
[12]	10 people 100 HS each	MFCC LBFC	GMM VQ
[14]	52 people 100 m / person	Multiple	SVM
[11]	10 people 20 HS each	Energy peaks	Euclidean distance
[9]	21 people 6 HS each 8 s / HS	MFCC, LDA, energy peaks	Euclidean distance
[8]	40 people 10 HS 10 s / HS	auto-corr. cross-corr. complex cepstrum	MSE kNN
[3]	40 people 2 HS each ~6 s / HS	MFCC FSR	Euclidean distance
[4]	165 people 2 HS each ~20 s / HS	LFCC FSR	GMM

Table 1. Recent approaches to heart-sound biometrics

A brief description of all those systems can be found in [6]. However, no performance comparison can be done since most of the systems use different performance metrics (EER, accuracy) and all of them use different, non publicly-available datasets.

4. Two Biometric Systems based on Heart Sounds

4.1. Algorithms for Heart-sounds Biometry

In this section we will describe some algorithms that are employed by both heart-sounds biometry systems, while in Sections 4.2 and 4.3 the algorithms that are specific to each system will be described.

4.1.1 Segmentation of S1 and S2

The algorithm that both systems use to extrapolate the S1 and S2 sounds from the heart sound signal is a variation of the algorithm introduced in [1]. The first step is splitting the signal in 4-seconds wide windows. Then, for each window, the algorithm finds the local maximum of the autocorrelation of the signal, that is the period P of the heart sound, and searches for other local maxima of the energy signal inside the window, at a distance $P \pm \epsilon$ from the first maximum.

These sounds are good candidates for being classified as S1 or S2, but still the class of the sound is not identified, and the set of sounds belonging to the other class must be found. So those sounds are zeroed out, and the algorithm is executed again in the window to find sounds belonging to the other class.

Finally, the algorithm classifies each set of sounds as S1 or S2 based on the fact that it is known fact that the distance from an S1 sound to the next S2 sound is always less than the distance from an S2 sound to the next S1 sound. This process is repeated for all the signal windows.

4.1.2 A Time-Domain Feature: the First-to-Second Ratio

Frequency (and quefrency) analysis is a feature extraction technique that heart-sounds biometry researchers have borrowed from the body of audio DSP techniques. The heart sound signal is something more than a generic audio file, it has a structure and probably this structure can be exploited to generate more meaningful features.

The first attempt to introducing a feature specific to heart sounds is done by the authors of [3], who in this work describe the First-to-Second Ratio (FSR). Intuitively, FSR can be thought as a representation of the average power ratio of S1 to S2. This feature comes from the experimental observation that some people have louder S1 sounds, while the others have louder S2 sounds.

The authors of [3] provide in this work a discussion on the implementation in the structural system, while details on the implementation in the statistical system are given in [4].

4.2. The Structural System

The system that we call "structural" was introduced in [1]; the name derives from the fact that it builds identity models as sequences of feature vectors, in opposition to the "statistical" system, described in Section 4.3, that builds models using statistical parameters learnt in the training phase.

The block diagram of the system is shown in Figure 1.

The first step of the identification process is the execution of the best subsequence detection algorithm described in [2], that splits the signal in 4-seconds wide windows and computes a quality index for each of the windows, retaining only the one with the best quality index. For each signal window i, the quality index is defined as:

$$DHS_{QI}(i) = \frac{1}{\sum_{k=1}^{4}\sum_{\substack{j=1\\ j\neq k}}^{4} d_{S1}(j,k) + \sum_{k=1}^{4}\sum_{\substack{j=1\\ j\neq k}}^{4} d_{S2}(j,k)} \tag{1}$$

Where d_{S1} and d_{S2} are cepstral distances computed between couples of heart sounds (j,k) defined as follows. Given two heart sounds signals X and Y, let $X_{S1}(i)$ (resp. $X_{S2}(i)$) be the 13-MFCC feature vector for the i-th S1 (resp. S2) sound of the X signal and Y_{S1} and Y_{S2} the analogous vectors for the Y signal. Then $d_{S1}(X,Y)$ and $d_{S2}(X,Y)$ are defined as follows:

$$d_{S1}(X,Y) = \frac{1}{N^2}\sum_{i,j=1}^{N} d_2(X_{S1}(i), Y_{S1}(j)) \tag{2}$$

$$d_{S2}(X,Y) = \frac{1}{N^2}\sum_{i,j=1}^{N} d_2(X_{S2}(i), Y_{S2}(j)) \tag{3}$$

where N is the number of cardiac cycles. Note that those two distances are also used in the matching phase.

After the best subsequence selection, the system uses the segmentation algorithm described in Section 4.1.1 to find the S1 and S2 tones in this subsequence. Then, feature extraction is carried on, using Mel-Frequency Cepstrum Coefficients (MFCC), and finally the matching score between the template M_X and the test sequence Y is computed as:

$$S(M_X, Y) = k_{FSR} \cdot \sqrt{d_{S1}(M_X,Y)^2 + d_{S2}(M_X,Y)^2} \tag{4}$$

In this equation, the k_{FSR} parameter is a multiplying factor depending on the average FSR of the two heart sound signals. For details on this parameter, see [3].

4.3. The Statistical System

The "statistical" system uses Gaussian Mixture Models (GMMs) [13] to represent identity models, and it is implemented using parts of the free and open source toolkit Alize/LIA_RAL [7], a speaker recognition framework, opportunely modified to deal with heart sound signals.

The system uses the GMM-UBM (Universal Background Model) recognition technique. This means that first it estimates the parameters of the world model λ_W using a random subset of the input signals, and then it derives the identity models λ_i using the Maximum A-Posteriori (MAP) algorithm. During the training phase, the system uses only the first of the two sequences per person that are present in the HSCT-11 database.

The block diagram of this system is shown in Figure 2.

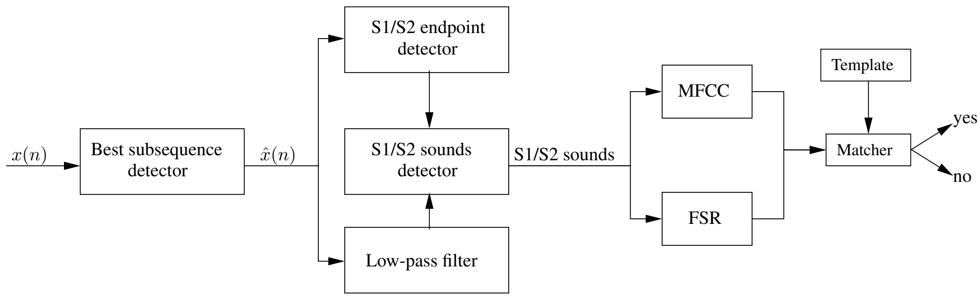

Figure 1. Block diagram of the identification process using the Structural system

The derivation of the parameters of GMM models is described in [5] and [4]. Namely, three feature sets were considered and the most effective one proved to be $16+16\Delta+E+\Delta E$ Linear Frequency Cepstral Coefficients (LFCC); four model sizes (128, 256, 512 and 1024 Gaussians per model) have been investigated and 256 Gaussians proved to give the best results. Finally the impact of the FSR was investigated, and it proved to be an useful feature because, all other things being equal, it allowed to obtain the same EER with half the number of Gaussians, thus speeding up the computation time by roughly a factor of 2.

5. Experimental Results

In this section we will present the HSCT-11 dataset, the evaluation protocol and the results of the performance evaluation of the two systems described in Section 4 done using this database.

5.1. The HSCT-11 Database

The HSCT-11 database is, to the best of our knowledge, the largest heart sounds database in terms of the number of people that contributed to it. It contains heart sounds acquired from 206 people, i.e. 157 male and 49 female.

From each person, we acquired two recordings; the average length of the sequences is 45 seconds, the minimum is 20 seconds and the maximum is 70 seconds. The two recordings were usually collected the same day, separated by a short break.

The sensor used for the acquisition is a ThinkLabs Rhythm Digital Electronic Stethoscope; the files were acquired using a sampling frequency of 11025 Hz and 16 bits per sample, and are stored using the WAVE format.

During the acquisition phase the person was sitting, in resting state, and the stethoscope was positioned near the pulmonary valve. So far, the database does not contain recordings of people after physical activity or of people with known cardiac illness, but we intend to expand it to cover also those cases.

The filenames encode the following metadata about the person:

- the first character encodes the sex of the person (M or F);
- the next 4 characters are the numeric ID of the person;
- the next character encodes the heart valve used for the auscultation (M: mitral, P: pulmonary, A: aortic, T: tricuspid); this database contains only sequences recorded near the pulmonary valve;
- the next character encodes whether the recording was done with the subject in resting condition (N) or after some light physical activity (C); so far the database contains only sequences recorded in resting condition;
- the next 3 characters encode the sequential number of the registration acquired from a given person; the first of these 3 characters is always the letter R.

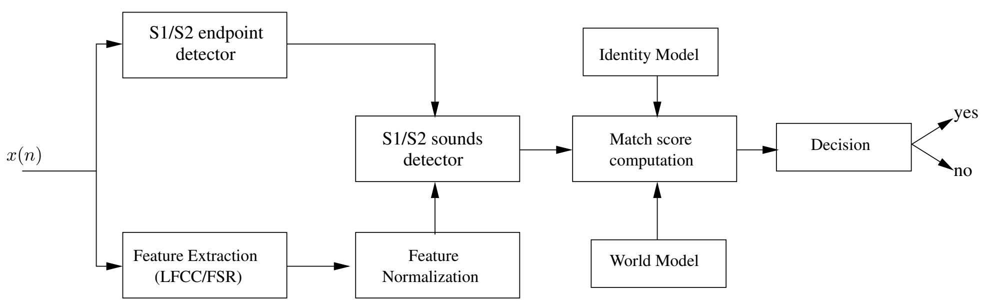

Figure 2. Block diagram of the identification process using the Statistical system

- the next 7 characters encode the date of the acquisition; the first one is always a letter D, the others represent the date in the format MMDDYY;

- the next 7 characters encode the birth date of the subject; the first one is always a letter N, the others represent the date in the format MMDDYY;

The letters between fields could have been avoided since the fields have a fixed length, but they have been inserted because they make it easier for human eyes to scan the filename and extract the required information. An example filename is: F7007NR01D290610N051077.wav.

Our database is freely available at the address *http://www.diit.unict.it/hsct11*

5.2. Evaluation Protocol

The comparison should be done in the following way: for each person, one sequence is used for the model training phase and one is used for the computation of matching scores.

Let X be a given person, X_a its first recording and X_b its second recording; also let D be the set of all the people in the database, and let $N = |D| = 206$ be the number of people in it. Let S be the matching function that, given an identity model and a recording gives a similarity score.

For each person, the database user should compute one genuine matching score, that is $S(M_X, X_b)$, and $N-1$ impostor matching scores $S(M_Y, X_b), \forall Y \in \{D \setminus X\}$. This will yield N genuine matching scores and $N \cdot (N-1)$ impostor matching scores.

5.3. Results

We evaluated the two systems described in Section 4 using the protocol presented in Section 5.2. As mentioned before, we carried on the evaluation using the HSCT-11 database.

Figure 3 shows the resulting Detection Error Tradeoff (DET) curves of the two systems, obtained plotting the False Match Rate (FMR) and False Non-Match Rate (FNMR) spanning the value of the detection threshold from the minimum to the maximum score value of each system. This figure also shows the Equal Error Rates (EER) of the two systems.

It is clear that the statistical system performs better than the structural one in every circumstance (low FNMR/high FMR - that means high usability, and low FMR/high FNMR - that means high security). This happens because the structural system was developed with a database that was much smaller and contained shorter sequences (~4-6 seconds). The algorithms used for the selection of the best subsequence are not enough for matching the performance of the statistical system.

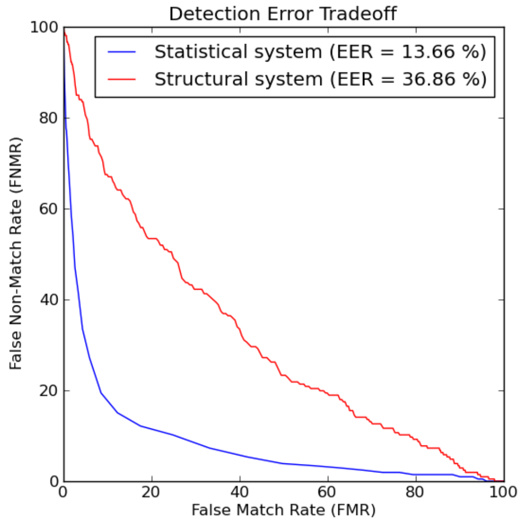

Figure 3. Detection Error Trade-off (DET) curves of the two systems

On the other hand, the performance of the statistical system is fairly good for a novel trait. The learning ability of Gaussian Mixture Models give them an advantage over the simple feature-based modeling technique used by the structural system. Also, the same system was tested in [4] using a subset of the HSCT-11 database containing heart sounds of 165 people, and in spite of a database size increase of around 20% the EER is substantially the same (13.66 % vs. 13.70%). This means that the approach seems to be robust to scaling to larger databases.

6. Conclusions and Future Work

In this paper, we proposed a performance evaluation study of recent heart sounds biometry systems based on the new open dataset HSCT-11.

From these results, it is clear that heart sounds biometry is a promising technique that is not yet ready for commercial usage. We believe that part of its immaturity also derives from the lack of a database to use as a common ground while testing the new approaches.

Some of the works described in Section 3 report lower values of EER, but the databases used in those tests, as can be seen from Table 1, are significantly smaller than the one used in this performance evaluation. So it is difficult - if not impossible - to compare the real difference in effectiveness of the approaches.

We hope that the release of an open dataset of heart sounds will encourage comparison of heart-sounds biometry systems on this common database, so that techniques can be objectively assessed and improved. We finally hope that other research groups will contribute to future releases of

HSCT-11 or will release their own datasets to the research community, so that future systems will be tested on open corpora of heart sounds, increasing the validity of each new scientific result obtained.

Acknowledgments

We are grateful to all the people that contributed their heart sounds to the HSCT-11 database and to all the people that helped in the acquisition of the heart sounds.

References

[1] F. Beritelli and S. Serrano. Biometric Identification based on Frequency Analysis of Cardiac Sounds. *IEEE Transactions on Information Forensics and Security*, 2(3):596–604, Sept. 2007.

[2] F. Beritelli and A. Spadaccini. Heart sounds quality analysis for automatic cardiac biometry applications. In *Proceedings of the 1st IEEE International Workshop on Information Forensics and Security*, Dec. 2009.

[3] F. Beritelli and A. Spadaccini. Human Identity Verification based on Mel Frequency Analysis of Digital Heart Sounds. In *Proceedings of the 16th International Conference on Digital Signal Processing*, July 2009.

[4] F. Beritelli and A. Spadaccini. An improved biometric identification system based on heart sounds and gaussian mixture models. In *Proceedings of the 2010 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, pages 31–35. IEEE, Sep 2010.

[5] F. Beritelli and A. Spadaccini. A statistical approach to biometric identity verification based on heart sounds. In *Proceedings of the Fourth International Conference on Emerging Security Information, Systems and Technologies (SECURWARE2010)*, pages 93–96. IEEE, Jul 2010.

[6] F. Beritelli and A. Spadaccini. *Human Identity Verification based on Heart Sounds: Recent Advances and Future Directions*, chapter 11, pages 217–234. June 2011.

[7] J.-F. Bonastre, N. Scheffer, D. Matrouf, C. Fredouille, A. Larcher, R. Preti, G. Pouchoulin, N. Evans, B. Fauve, and J. Mason. Alize/Spkdet: a state-of-the-art open source software for speaker recognition.

[8] N. El-Bendary, H. Al-Qaheri, H. M. Zawbaa, M. Hamed, A. E. Hassanien, Q. Zhao, and A. Abraham. Hsas: Heart sound authentication system. In *Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress on*, pages 351 –356, 2010.

[9] S. Fatemian, F. Agrafioti, and D. Hatzinakos. Heartid: Cardiac biometric recognition. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1 –5, 2010.

[10] A. K. Jain, A. A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(2):4–20, Jan 2004.

[11] J. Jasper and K. Othman. Feature extraction for human identification based on envelogram signal analysis of cardiac sounds in time-frequency domain. In *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, volume 2, pages V2–228 –V2–233, 2010.

[12] K. Phua, J. Chen, T. H. Dat, and L. Shue. Heart sound as a biometric. *Pattern Recognition*, 41(3):906–919, Mar 2008.

[13] D. A. Reynolds and R. C. Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3:72–83, 1995.

[14] D. H. Tran, Y. R. Leng, and H. Li. Feature integration for heart sound biometrics. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 1714 –1717, 2010.

[15] T. Ye-wei, S. Xia, Z. Hui-xiang, and W. Wei. A biometric identification system based on heart sound signal. In *Human System Interactions (HSI), 2010 3rd Conference on*, pages 67 –75, May 2010.