Human Identity Verification based on Heart Sounds: Recent Advances and Future Directions

Francesco Beritelli and Andrea Spadaccini
Dipartimento di Ingegneria Elettrica, Elettronica ed Informatica (DIEEI)
University of Catania
Italy

1. Introduction

Identity verification is an increasingly important process in our daily lives. Whether we need to use our own equipment or to prove our identity to third parties in order to use services or gain access to physical places, we are constantly required to declare our identity and prove our claim.
Traditional authentication methods fall into two categories: proving that you know something (i.e., password-based authentication) and proving that you own something (i.e., token-based authentication).
These methods connect the identity with an alternate and less rich representation, for instance a password, that can be lost, stolen, or shared.
A solution to these problems comes from biometric recognition systems. Biometrics offers a natural solution to the authentication problem, as it contributes to the construction of systems that can recognize people by the analysis of their anatomical and/or behavioral characteristics. With biometric systems, the representation of the identity is something that is directly derived from the subject, therefore it has properties that a surrogate representation, like a password or a token, simply cannot have (Jain et al. (2006; 2004); Prabhakar et al. (2003)).
The strength of a biometric system is determined mainly by the trait that is used to verify the identity. Plenty of biometric traits have been studied and some of them, like fingerprint, iris and face, are nowadays used in widely deployed systems.
Today, one of the most important research directions in the field of biometrics is the characterization of novel biometric traits that can be used in conjunction with other traits, to limit their shortcomings or to enhance their performance.
The aim of this chapter is to introduce the reader to the usage of heart sounds for biometric recognition, describing the strengths and the weaknesses of this novel trait and analyzing in detail the methods developed so far and their performance.
The usage of heart sounds as physiological biometric traits was first introduced in Beritelli & Serrano (2007), in which the authors proposed and started exploring this idea. Their system is based on the frequency analysis, by means of the Chirp z-Transform (CZT), of the sounds produced by the heart during the closure of the mitral tricuspid valve and during the closure of the aortic pulmonary valve. These sounds, called S1 and S2, are extracted from the input signal using a segmentation algorithm. The authors build the identity templates

using feature vectors and test if the identity claim is true by computing the Euclidean distance between the stored template and the features extracted during the identity verification phase.
In Phua et al. (2008), the authors describe a different approach to heart-sounds biometry. Instead of doing a structural analysis of the input signal, they use the whole sequences, feeding them to two recognizers built using Vector Quantization and Gaussian Mixture Models; the latter proves to be the most performant system.
In Beritelli & Spadaccini (2009a;b), the authors further develop the system described in Beritelli & Serrano (2007), evaluating its performance on a larger database, choosing a more suitable feature set (Linear Frequency Cepstrum Coefficients, LFCC), adding a time-domain feature specific for heart sounds, called First-to-Second Ratio (FSR) and adding a quality-based data selection algorithm.
In Beritelli & Spadaccini (2010a;b), the authors take an alternative approach to the problem, building a system that leverages statistical modelling using Gaussian Mixture Models. This technique is different from Phua et al. (2008) in many ways, most notably the segmentation of the heart sounds, the database, the usage of features specific to heart sounds and the statistical engine. This system proved to yield good performance in spite of a larger database, and the final Equal Error Rate (EER) obtained using this technique is 13.70 % over a database of 165 people, containing two heart sequences per person, each lasting from 20 to 70 seconds.
This chapter is structured as follows: in Section 2, we describe in detail the usage of heart sounds for biometric identification, comparing them to other biometric traits, briefly explaining how the human cardio-circulatory system works and produces heart sounds and how they can be processed; in Section 3 we present a survey of recent works on heart-sounds biometry by other research groups; in Section 4 we describe in detail the structural approach; in Section 5 we describe the statistical approach; in Section 6 we compare the performance of the two methods on a common database, describing both the performance metrics and the heart sounds database used for the evaluation; finally, in Section 7 we present our conclusions, and highlight current issues of this method and suggest the directions for the future research.

2. Biometric recognition using heart sounds

Biometric recognition is the process of inferring the identity of a person via quantitative analysis of one or more traits, that can be derived either directly from a person's body (physiological traits) or from one's behaviour (behavioural traits).
Speaking of physiological traits, almost all the parts of the body can already be used for the identification process (Jain et al. (2008)): eyes (iris and retina), face, hand (shape, veins, palmprint, fingerprints), ears, teeth etc.
In this chapter, we will focus on an organ that is of fundamental importance for our life: the heart.
The heart is involved in the production of two biological signals, the Electrocardiograph (ECG) and the Phonocardiogram (PCG). The first is a signal derived from the electrical activity of the organ, while the latter is a recording of the sounds that are produced during its activity (heart sounds).
While both signals have been used as biometric traits (see Biel et al. (2001) for ECG-based biometry), this chapter will focus on hearts-sounds biometry.

2.1 Comparison to other biometric traits

The paper Jain et al. (2004) presents a classification of available biometric traits with respect to 7 qualities that, according to the authors, a trait should possess:

- **Universality**: each person should possess it;
- **Distinctiveness**: it should be helpful in the distinction between any two people;
- **Permanence**: it should not change over time;
- **Collectability**: it should be quantitatively measurable;
- **Performance**: biometric systems that use it should be reasonably performant, with respect to speed, accuracy and computational requirements;
- **Acceptability**: the users of the biometric system should see the usage of the trait as a natural and trustable thing to do in order to authenticate;
- **Circumvention**: the system should be robust to malicious identification attempts.

Each trait is evaluated with respect to each of these qualities using 3 possible qualifiers: H (high), M (medium), L (low).
We added to the original table a row with our subjective evaluation of heart-sounds biometry with respect to the qualities described above, in order to compare this new technique with other more established traits. The updated table is reproduced in Table 1.
The reasoning behind each of our subjective evaluations of the qualities of heart sounds is as follows:

- **High Universality**: a working heart is a *conditio sine qua non* for human life;
- **Medium Distinctiveness**: the actual systems' performance is still far from the most discriminating traits, and the tests are conducted using small databases; the discriminative power of heart sounds still must be demonstrated;
- **Low Permanence**: although to the best of our knowledge no studies have been conducted in this field, we perceive that heart sounds can change their properties over time, so their accuracy over extended time spans must be evaluated;
- **Low Collectability**: the collection of heart sounds is not an immediate process, and electronic stethoscopes must be placed in well-defined positions on the chest to get a high-quality signal;
- **Low Performance**: most of the techniques used for heart-sounds biometry are computationally intensive and, as said before, the accuracy still needs to be improved;
- **Medium Acceptability**: heart sounds are probably identified as unique and trustable by people, but they might be unwilling to use them in daily authentication tasks;
- **Low Circumvention**: 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.

Of course, heart-sounds biometry is a new technique, and some of its drawbacks probably will be addressed and resolved in future research work.

Biometric identifier	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Circumvention
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial thermogram	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand geometry	M	M	M	H	M	M	M
Hand vein	M	M	M	M	M	M	L
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Voice	M	L	M	L	L	M	H
Heart sounds	H	M	L	L	L	M	L

Table 1. Comparison between biometric traits as in Jain et al. (2004) and heart sounds

2.2 Physiology and structure of heart sounds

The heart sound signal is a complex, non-stationary and quasi-periodic signal that is produced by the heart during its continuous pumping work (Sabarimalai Manikandan & Soman (2010)). It is composed by several smaller sounds, each associated with a specific event in the working cycle of the heart.
Heart sounds fall in two categories:

- **primary sounds**, produced by the closure of the heart valves;
- **other sounds**, produced by the blood flowing in the heart or by pathologies;

The primary sounds are S1 and S2. The first sound, S1, is caused by the closure of the tricuspid and mitral valves, while the second sound, S2, is caused by the closure of the aortic and pulmonary valves.
Among the other sounds, there are the S3 and S4 sounds, that are quieter and rarer than S1 and S2, and murmurs, that are high-frequency noises.
In our systems, we only use the primary sounds because they are the two loudest sounds and they are the only ones that a heart always produces, even in pathological conditions. We separate them from the rest of the heart sound signal using the algorithm described in Section 2.3.1.

2.3 Processing heart sounds

Heart sounds are monodimensional signals, and can be processed, to some extent, with techniques known to work on other monodimensional signals, like audio signals. Those

techniques then need to be refined taking into account the peculiarities of the signal, its structure and components.
In this section we will describe an algorithm used to separate the S1 and S2 sounds from the rest of the heart sound signal (2.3.1) and three algorithms used for feature extraction (2.3.2, 2.3.3, 2.3.4), that is the process of transforming the original heart sound signal into a more compact, and possibly more meaningful, representation. We will briefly discuss two algorithms that work in the frequency domain, and one in the time domain.

2.3.1 Segmentation

In this section we describe a variation of the algorithm that was employed in (Beritelli & Serrano (2007)) to separate the S1 and S2 tones from the rest of the heart sound signal, improved to deal with long heart sounds.
Such a separation is done because we believe that the S1 and S2 tones are as important to heart sounds as the vowels are to the voice signal. They are stationary in the short term and they convey significant biometric information, that is then processed by feature extraction algorithms.
A simple energy-based approach can not be used because the signal can contain impulsive noise that could be mistaken for a significant sound.
The first step of the algorithm is searching the frame with the highest energy, that is called SX1. At this stage, we do not know if we found an S1 or an S2 sound.
Then, in order to estimate the frequency of the heart beat, and therefore the period P of the signal, the maximum value of the autocorrelation function is computed. Low-frequency components are ignored by searching only over the portion of autocorrelation after the first minimum.
The algorithm then searches other maxima to the left and to the right of SX1, moving by a number P of frames in each direction and searching for local maxima in a window of the energy signal in order to take into account small fluctuations of the heart rate. After each maximum is selected, a constant-width window is applied to select a portion of the signal.
After having completed the search that starts from SX1, all the corresponding frames in the original signal are zeroed out, and the procedure is repeated to find a new maximum-energy frame, called SX2, and the other peaks are found in the same way.
Finally, the positions of SX1 and SX2 are compared, and the algorithm then decides if SX1, and all the frames found starting from it, must be classified as S1 or S2; the remaining identified frames are classified accordingly.
The nature of this algorithm requires that it work on short sequences, 4 to 6 seconds long, because as the sequence gets longer the periodicity of the sequence fades away due to noise and variations of the heart rate.
To overcome this problem, the signal is split into 4-seconds wide windows and the algorithm is applied to each window. The resulting sets of heart sounds endpoint are then joined into a single set.

2.3.2 The chirp z-transform

The Chirp z-Transform (CZT) is an algorithm for the computation of the z-Transform of sampled signals that offers some additional flexibility to the Fast Fourier Transform (FFT) algorithm.

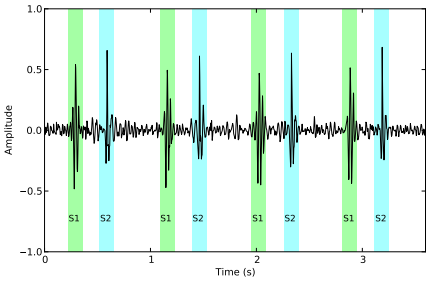

Fig. 1. Example of S1 and S2 detection

The main advantage of the CZT exploited in the analysis of heart sounds is the fact that it allows high-resolution analysis of narrow frequency bands, offering higher resolution than the FFT.
For more details on the CZT, please refer to Rabiner et al. (1969)

2.3.3 Cepstral analysis

Mel-Frequency Cepstrum Coefficients (MFCC) are one of the most widespread parametric representation of audio signals (Davis & Mermelstein (1980)).
The basic idea of MFCC is the extraction of cepstrum coefficients using a non-linearly spaced filterbank; the filterbank is instead spaced according to the Mel Scale: filters are linearly spaced up to 1 kHz, and then are logarithmically spaced, decreasing detail as the frequency increases.
This scale is useful because it takes into account the way we perceive sounds.
The relation between the Mel frequency $\hat{f}_{mel}$ and the linear frequency f_{lin} is the following:

$$\hat{f}_{mel} = 2595 \cdot \log_{10}\left(\frac{1 + f_{lin}}{700}\right) \tag{1}$$

Some heart-sound biometry systems use MFCC, while others use a linearly-spaced filterbank.
The first step of the algorithm is to compute the FFT of the input signal; the spectrum is then feeded to the filterbank, and the i-th cepstrum coefficient is computed using the following formula:

$$C_i = \sum_{k=1}^{K} X_k \cdot \cos\left(i \cdot \left(k - \frac{1}{2}\right) \cdot \frac{\pi}{K}\right) i = 0, ..., M \tag{2}$$

where K is the number of filters in the filterbank, X_k is the log-energy output of the k-th filter and M is the number of coefficients that must be computed.
Many parameters have to be chosen when computing cepstrum coefficients. Among them: the bandwidth and the scale of the filterbank (Mel vs. linear), the number and spectral width of filters, the number of coefficients.
In addition to this, differential cepstrum coefficients, tipically denoted using a Δ (first order) or $\Delta\Delta$ (second order), can be computed and used.
Figure 2 shows an example of three S1 sounds and the relative MFCC spectrograms; the first two (a, b) belong to the same person, while the third (c) belongs to a different person.

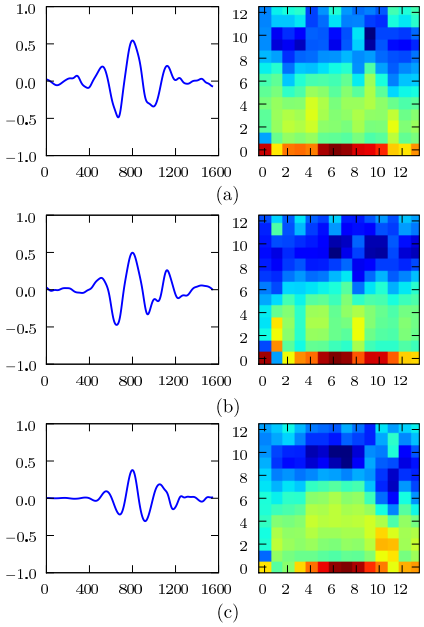

Fig. 2. Example of waveforms and MFCC spectrograms of S1 sounds

2.3.4 The First-to-Second Ratio (FSR)

In addition to standard feature extraction techniques, it would be desirable to develop ad-hoc features for the heart sound, as it is not a simple audio sequence but has specific properties that could be exploited to develop features with additional discriminative power.
This is why we propose a time-domain feature called First-to-Second Ratio (FSR). Intuitively, the FSR represents the power ratio of the first heart sound (S1) to the second heart sound (S2). During our work, we observed that some people tend to have an S1 sound that is louder than S2, while in others this balance is inverted. We try to represent this diversity using our new feature.
The implementation of the feature is different in the two biometric systems that we described in this chapter, and a discussion of the two algorithms can be found in 4.4 and 5.4.

3. Review of related works

In the last years, different research groups have been studying the possibility of using heart sounds for biometric recognition. In this section, we will briefly describe their methods.
In Table 2 we summarized the main characteristics of the works that will be analyzed in this section, 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.

We chose not to represent performance in this table for two reasons: first, most papers do not adopt the same performance metric, so it would be difficult to compare them; second, the database and the approach used are quite different one from another, so it would not be a fair comparison.

Paper	Database	Features	Classification
Phua et al. (2008)	10 people 100 HS each	MFCC LBFC	GMM VQ
Tran et al. (2010)	52 people 100m each	Multiple	SVM
Jasper & Othman (2010)	10 people 20 HS each	Energy peaks	Euclidean distance
Fatemian et al. (2010)	21 people 6 HS each 8 seconds per HS	MFCC, LDA, energy peaks	Euclidean distance
El-Bendary et al. (2010)	40 people 10 HS 10 seconds per HS	autocorrelation cross-correlation complex cepstrum	MSE kNN

Table 2. Comparison of recent works about heart-sound biometrics

In the rest of the section, we will briefly review each of these papers.
Phua et al. (2008) was one of the first works in the field of heart-sounds biometry. In this paper, the authors first do a quick exploration of the feasibility of using heart sounds as a biometric trait, by recording a test database composed of 128 people, using 1-minute heart sounds and splitting the same signal into a train and a testing sequence. Having obtained good recognition performance using the HTK Speech Recognition toolkit, they do a deeper test using a database recorded from 10 people and containing 100 sounds for each person, investigating the performance of the system using different feature extraction algorithms (MFCC, Linear Frequency Band Cepstra (LFBC)), different classification schemes (Vector Quantization (VQ) and Gaussian Mixture Models (GMM)) and investigating the impact of the frame size and of the training/test length. After testing many combinations of those parameters, they conclude that, on their database, the most performing system is composed of LFBC features (60 cepstra + log-energy + 256ms frames with no overlap), GMM-4 classification, 30s of training/test length.
The authors of Tran et al. (2010), one of which worked on Phua et al. (2008), take the idea of finding a good and representative feature set for heart sounds even further, exploring 7 sets of features: temporal shape, spectral shape, cepstral coefficientrs, harmonic features, rhythmic features, cardiac features and the GMM supervector. They then feed all those features to a feature selection method called RFE-SVM and use two feature selection strategies (optimal and sub-optimal) to find the best set of features among the ones they considered. The tests

were conducted on a database of 52 people and the results, expressed in terms of Equal Error Rate (EER), are better for the automatically selected feature sets with respect to the EERs computed over each individual feature set.
In Jasper & Othman (2010), the authors describe an experimental system where the signal is first downsampled from 11025 Hz to 2205 Hz; then it is processed using the Discrete Wavelet Transform, using the Daubechies-6 wavelet, and the D4 and D5 subbands (34 to 138 Hz) are then selected for further processing. After a normalization and framing step, the authors then extract from the signal some energy parameters, and they find that, among the ones considered, the Shannon energy envelogram is the feature that gives the best performance on their database of 10 people.
The authors of Fatemian et al. (2010) do not propose a pure-PCG approach, but they rather investigate the usage of both the ECG and PCG for biometric recognition. In this short summary, we will focus only on the part of their work that is related to PCG. The heart sounds are processed using the Daubechies-5 wavelet, up to the 5th scale, and retaining only coefficients from the 3rd, 4th and 5th scales. They then use two energy thresholds (low and high), to select which coefficients should be used for further stages. The remaining frames are then processed using the Short-Term Fourier Transform (STFT), the Mel-Frequency filterbank and Linear Discriminant Analysis (LDA) for dimensionality reduction. The decision is made using the Euclidean distance from the feature vector obtained in this way and the template stored in the database. They test the PCG-based system on a database of 21 people, and their combined PCG-ECG systems has better performance.
The authors of El-Bendary et al. (2010) filter the signal using the DWT; then they extract different kinds of features: auto-correlation, cross-correlation and cepstra. They then test the identities of people in their database, that is composed by 40 people, using two classifiers: Mean Square Error (MSE) and k-Nearest Neighbor (kNN). On their database, the kNN classifier performs better than the MSE one.

4. The structural approach to heart-sounds biometry

The first system that we describe in depth was introduced in Beritelli & Serrano (2007); it was designed to work with short heart sounds, 4 to 6 seconds long and thus containing at least four cardiac cycles (S1-S2).
The restriction on the length of the heart sound was removed in Beritelli & Spadaccini (2009a), that introduced the quality-based best subsequence selection algorithm, described in 4.1.
We call this system “structural” because the identity templates are stored as feature vectors, in opposition to the “statistical” approach, that does not directly keep the feature vectors but instead it represents identities via statistical parameters inferred in the learning phase.
Figure 3 contains the block diagram of the system. Each of the steps will be described in the following sections.

4.1 The best subsequence selection algorithm

The fact that the segmentation and matching algorithms of the original system were designed to work on short sequences was a strong constraint for the system. It was required that a human operator selected a portion of the input signal based on some subjective assumptions. It was clearly a flaw that needed to be addressed in further versions of the system.

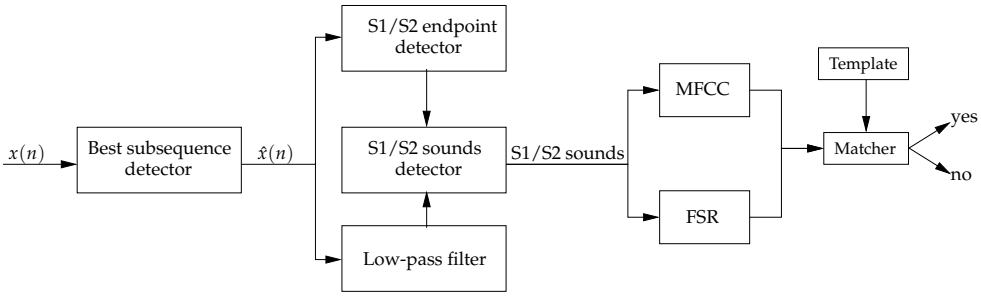

Fig. 3. Block diagram of the proposed cardiac biometry system

To resolve this issue, the authors developed a quality-based subsequence selection algorithm, based on the definition of a quality index $DHS_{QI}(i)$ for each contiguous subsequence i of the input signal.
The quality index is based on a cepstral similarity criterion: the selected subsequence is the one for which the cepstral distance of the tones is the lowest possible. So, for a given subsequence i, the quality index is defined as:

$$DHS_{QI}(i) = \frac{1}{\sum_{k=1}^{4}\sum_{\substack{j=1 \\ j\neq k}}^{4} d_{S1}(j,k) + \sum_{k=1}^{4}\sum_{\substack{j=1 \\ j\neq k}}^{4} d_{S2}(j,k)} \tag{3}$$

Where d_{S1} and d_{S2} are the cepstral distances defined in 4.5.
The subsequence $\bar{i}$ with the maximum value of $DHS_{QI}(\bar{i})$ is then selected as the best one and retained for further processing, while the rest of the input signal is discarded.

4.2 Filtering and segmentation

After the best subsequence selection, the signal is then given in input to the heart sound endpoint detection algorithm described in 2.3.1.
The endpoints that it finds are then used to extract the relevant portions of the signal over a version of the heart sound signal that was previously filtered using a low-pass filter, which removed the high-frequency extraneous components.

4.3 Feature extraction

The heart sounds are then passed to the feature extraction module, that computes the cepstral features according to the algorithm described in 2.3.
This system uses $M = 12$ MFCC coefficients, with the addition of a 13-th coefficient computed using an $i = 0$ value in Equation 2, that is the log-energy of the analyzed sound.

4.4 Computation of the First-to-Second Ratio

For each input signal, the system computes the FSR according to the following algorithm.
Let N be the number of complete S1-S2 cardiac cycles in the signal. Let P_{S1_i} (resp. P_{S2_i}) be the power of the i-th S1 (resp. S2) sound.
We can then define $\overline{P_{S1}}$ and $\overline{P_{S2}}$, the average powers of S1 and S2 heart sounds:

$$\overline{P_{S1}} = \frac{1}{N}\sum_{i=1}^{N} P_{S1_i} \tag{4}$$

$$\overline{P_{S2}} = \frac{1}{N}\sum_{i=1}^{N} P_{S2_i} \tag{5}$$

Using these definitions, we can then define the First-to-Second Ration of a given heart sound signal as:

$$FSR = \frac{\overline{P_{S1}}}{\overline{P_{S2}}} \tag{6}$$

For two given DHS sequences x_1 and x_2, we define the FSR distance as:

$$d_{FSR}(x_1, x_2) = |FSR_{dB}(x_1) - FSR_{dB}(x_2)| \tag{7}$$

4.5 Matching and identity verification

The crucial point of identity verification is the computation of the distance between the feature set that represents the input signal and the template associated with the identity claimed in the acquisition phase by the person that is trying to be authenticated by the system.
This system employs two kinds of distance: the first in the cepstral domain and the second using the FSR.
MFCC are compared using the Euclidean metric (d_2). Given two heart sound signals X and Y, let $X_{S1}(i)$ (resp. $X_{S2}(i)$) be the 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 the cepstral distances between X and Y can be defined as follows:

$$d_{S1}(X, Y) = \frac{1}{N^2}\sum_{i,j=1}^{N} d_2(X_{S1}(i), Y_{S1}(j)) \tag{8}$$

$$d_{S2}(X, Y) = \frac{1}{N^2}\sum_{i,j=1}^{N} d_2(X_{S2}(i), Y_{S2}(j)) \tag{9}$$

Now let us take into account the FSR. Starting from the d_{FSR} as defined in Equation 7, we wanted this distance to act like an amplifying factor for the cepstral distance, making the distance bigger when it has an high value while not changing the distance for low values.
We then normalized the values of d_{FSR} between 0 and 1 ($d_{FSR_{norm}}$), we chose a threshold of activation of the FSR (th_FSR) and we defined defined k_{FSR}, an amplifying factor used in the matching phase, as follows:

$$k_{FSR} = \max\left(1, \frac{d_{FSR_{norm}}}{th_{FSR}}\right) \tag{10}$$

In this way, if the normalized FSR distance is lower than th_{FSR} it has no effect on the final score, but if it is larger, it will increase the cepstral distance.
Finally, the distance between X and Y can be computed as follows:

$$d(X, Y) = k_{FSR} \cdot \sqrt{d_{S1}(X, Y)^2 + d_{S2}(X, Y)^2} \tag{11}$$

5. The statistical approach to heart-sounds biometry

In opposition to the system analyzed in Section 4, the one that will be described in this section is based on a learning process that does not directly take advantage of the features extracted from the heart sounds, but instead uses them to infer a statistical model of the identity and makes a decision computing the probability that the input signal belongs to the person whose identity was claimed in the identity verification process.

5.1 Gaussian Mixture Models

Gaussian Mixture Models (GMM) are a powerful statistical tool used for the estimation of multidimensional probability density representation and estimation (Reynolds & Rose (1995)).
A GMM λ is a weighted sum of N Gaussian probability densities:

$$p(\boldsymbol{x}|\lambda) = \sum_{i=1}^{N} w_i p_i(\boldsymbol{x}) \tag{12}$$

where $\boldsymbol{x}$ is a D-dimensional data vector, whose probability is being estimated, and w_i is the weight of the i-th probability density, that is defined as:

$$p_i(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|}} e^{-\frac{1}{2}(\boldsymbol{x}-\mu_i)'\Sigma_i(\boldsymbol{x}-\mu_i)}$$

The parameters of p_i are μ_i ($\in \mathbb{R}^D$) and Σ_i ($\in \mathbb{R}^{D\times D}$), that together with w_i ($\in \mathbb{R}^N$) form the set of values that represent the GMM:

$$\lambda = \{w_i, \mu_i, \Sigma_i\} \tag{13}$$

Those parameters of the model are learned in the training phase using the Expectation-Maximization algorithm (McLachlan & Krishnan (1997)), using as input data the feature vectors extracted from the heart sounds.

5.2 The GMM/UBM method

The problem of verifying whether an input heart sound signal s belongs to a stated identity I is equivalent to a hypothesis test between two hypotheses:

$$H_0 : s \text{ belongs to } I$$
$$H_1 : s \text{ does not belong to } I$$

This decision can be taken using a likelihood test:

$$S(s, I) = \frac{p(s|H_0)}{p(s|H_1)} \begin{cases} \geq \theta \text{ accept } H_0 \\ < \theta \text{ reject } H_0 \end{cases} \tag{14}$$

where θ is the decision threshold, a fundamental system parameter that is chosen in the design phase.
The probability $p(s|H_0)$, in our system, computed using Gaussian Mixture Models.

The input signal is converted by the front-end algorithms to a set of K feature vectors, each of dimension D, so:

$$p(s|H_0) = \prod_{j=1}^{K} p(x_j|\lambda_I) \tag{15}$$

In Equation 14, the $p(s|H_1)$ is still missing. In the GMM/UBM framework (Reynolds et al. (2000)), this probability is modelled by building a model trained with a set of identities that represent the demographic variability of the people that might use the system. This model is called Universal Background Model (UBM).
The UBM is created during the system design, and is subsequently used every time the system must compute a matching score.
The final score of the identity verification process, expressed in terms of log-likelihood ratio, is

$$\Lambda(s) = \log S(s, I) = \log p(s|\lambda_I) - \log p(s|\lambda_W) \tag{16}$$

5.3 Front-end processing

Each time the system gets an input file, whether for training a model or for identity verification, it goes through some common steps.
First, heart sounds segmentation is carried on, using the algorithm described in Section 2.3.1.
Then, cepstral features are extracted using a tool called *sfbcep*, part of the SPro suite (Gravier (2003)). Finally, the FSR, computed as described in Section 5.4, is appended to each feature vector.

5.4 Application of the First-to-Second Ratio

The FSR, as first defined in Section 4.4, is a sequence-wise feature, i.e., it is defined for the whole input signal. It is then used in the matching phase to modify the resulting score.
In the context of the statistical approach, it seemed more appropriate to just append the FSR to the feature vector computed from each frame in the feature extraction phase, and then let the GMM algorithms generalize this knowledge.
To do this, we split the input heart sound signal in 5-second windows and we compute an average FSR ($\overline{FSR}$) for each signal window. It is then appended to each feature vector computed from frames inside the window.

5.5 The experimental framework

The experimental set-up created for the evaluation of this technique was implemented using some tools provided by ALIZE/SpkDet , an open source toolkit for speaker recognition developed by the ELISA consortium between 2004 and 2008 (Bonastre et al. (n.d.)).
The adaptation of parts of a system designed for speaker recognition to a different problem was possible because the toolkit is sufficiently general and flexible, and because the features used for heart-sounds biometry are similar to the ones used for speaker recognition, as outlined in Section 2.3.
During the world training phase, the system estimates the parameters of the world model λ_W using a randomly selected subset of the input signals.
The identity models λ_i are then derived from the world model W using the Maximum A-Posteriori (MAP) algorithm.
During identity verification, the matching score is computed using Equation 16, and the final decision is taken comparing the score to a threshold (θ), as described in Equation 14

5.6 Optimization of the method

During the development of the system, some parameters have been tuned in order to get the best performance. Namely, three different cepstral feature sets have been considered in (Beritelli & Spadaccini (2010b)):

- $16 + 16\,\Delta + E + \Delta E$
- $16 + 16\,\Delta + 16\,\Delta\Delta$
- $19 + 19\,\Delta + E + \Delta E$

However, the first of these sets proved to be the most effective
In (Beritelli & Spadaccini (2010a)) the impact of the FSR and of the number of Gaussian densities in the mixtures was studied. Four different model sizes (128, 256, 512, 1024) were tested, with and without FSR, and the best combination of those parameters, on our database, is 256 Gaussians with FSR.

6. Performance evaluation

In this section, we will compare the performance of the two systems described in Section 4 and 5 using a common heart sounds database, that will be further described in Section 6.1.

6.1 Heart sounds database

One of the drawbacks of this biometric trait is the absence of large enough heart sound databases, that are needed for the validation of biometric systems. To overcome this problem, we are building a heart sounds database suitable for identity verification performance evaluation.
Currently, there are 206 people in the database, 157 male and 49 female; for each person, there are two separate recordings, each lasting from 20 to 70 seconds; the average length of the recordings is 45 seconds. The heart sounds have been acquired using a Thinklabs Rhythm Digital Electronic Stethoscope, connected to a computer via an audio card. The sounds have been converted to the Wave audio format, using 16 bit per second and at a rate of 11025 Hz.
One of the two recordings available for each personused to build the models, while the other is used for the computation of matching scores.

6.2 Metrics for performance evaluation

A biometric identity verification system can be seen as a binary classifier.
Binary classification systems work by comparing matching scores to a threshold; their accuracy is closely linked with the choice of the threshold, which must be selected according to the context of the system.
There are two possible errors that a binary classifier can make:

- **False Match (Type I Error):** accept an identity claim even if the template does not match with the model;
- **False Non-Match (Type II Error):** reject an identity claim even if the template matches with the model

The importance of errors depends on the context in which the biometric system operates; for instance, in a high-security environment, a Type I error can be critical, while Type II errors could be tolerated.

When evaluating the performance of a biometric system, however, we need to take a threshold-independent approach, because we cannot know its applications in advance. A common performance measure is the Equal Error Rate (EER) (Jain et al. (2008)), defined as the error rate at which the False Match Rate (FMR) is equal to the False Non-Match Rate (FNMR). A finer evaluation of biometric systems can be done by plotting the Detection Error Tradeoff (DET) curve, that is the plot of FMR against FNMR. This allows to study their performance when a low FNMR or FMR is imposed to the system.
The DET curve represents the trade-off between security and usability. A system with low FMR is a highly secure one but will lead to more non-matches, and can require the user to try the authentication step more times; a system with low FNMR will be more tolerant and permissive, but will make more false match errors, thus letting more unauthorized users to get a positive match. The choice between the two setups, and between all the intermediate security levels, is strictly application-dependent.

6.3 Results

The performance of our two systems has been computed over the heart sounds database, and the results are reported in Table 3.

System	EER (%)
Structural	36.86
Statistical	**13.66**

Table 3. Performance evaluation of the two heart-sounds biometry systems

The huge difference in the performance of the two systems reflects the fact that the first one is not being actively developed since 2009, and it was designed to work on small databases, while the second has already proved to work well on larger databases.
It is important to highlight that, in spite of a 25% increment of the size of the database, the error rate remained almost constant with respect to the last evaluation of the system, in which a test over a 165 people database yielded a 13.70% EER.
Figure 4 shows the Detection Error Trade-off (DET) curves of the two systems. As stated before, a DET curve shows how the analyzed system performs in terms of false matches/false non-matches as the system threshold is changed.
In both cases, fixing a false match (resp. false non-match) rate, the system that performs better is the one with the lowest false non-match (resp. false match) rate.
Looking at Figure 4, it is easy to understand that the statistical system performs better in both high-security (e.g., FMR = 1-2%) and low-security (e.g., FNMR = 1-2%) setups.
We can therefore conclude that the statistical approach is definitely more promising that the structural one, at least with the current algorithms and using the database described in 6.1..

7. Conclusions

In this chapter, we presented a novel biometric identification technique that is based on heart sounds.
After introducing the advantages and shortcomings of this biometric trait with respect to other traits, we explained how our body produces heart sounds, and the algorithms used to process them.

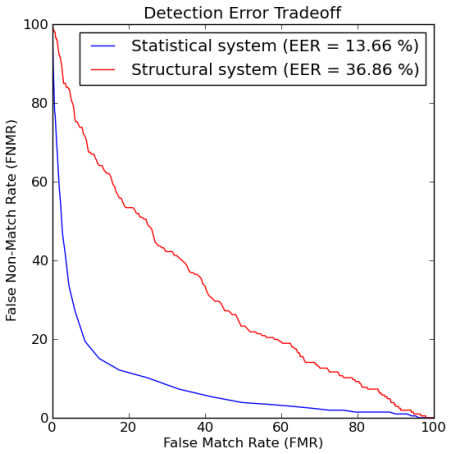

Fig. 4. Detection Error Tradeoff (DET) curves of the two systems

A survey of recent works on this field written by other research groups has been presented, showing that there has been a recent increase of interest of the research community in this novel trait.
Then, we described the two systems that we built for biometric identification based on heart sounds, one using a structural approach and another leveraging Gaussian Mixture Models. We compared their performance over a database containing more than 200 people, concluding that the statistical system performs better.

7.1 Future directions

As this chapter has shown, heart sounds biometry is a promising research topic in the field of novel biometric traits.
So far, the academic community has produced several works on this topic, but most of them share the problem that the evaluation is carried on over small databases, making the results obtained difficult to generalize.
We feel that the community should start a joint effort for the development of systems and algorithms for heart-sounds biometry, at least creating a common database to be used for the evaluation of different research systems over a shared dataset that will make possible to compare their performance in order to refine them and, over time, develop techniques that might be deployed in real-world scenarios.
As larger databases of heart sounds become available to the scientific community, there are some issues that need to be addressed in future research.
First of all, the identification performance should be kept low even for larger databases. This means that the matching algorithms will be fine-tuned and a suitable feature set will be identified, probably containing both elements from the frequency domain and the time domain.
Next, the mid-term and long-term reliability of heart sounds will be assessed, analyzing how their biometric properties change as time goes by. Additionally, the impact of cardiac diseases on the identification performance will be assessed.

Finally, when the algorithms will be more mature and several independent scientific evaluations will have given positive feedback on the idea, some practical issues like computational efficiency will be tackled, and possibly ad-hoc sensors with embedded matching algorithms will be developed, thus making heart-sounds biometry a suitable alternative to the mainstream biometric traits.

8. References

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

Beritelli, F. & Spadaccini, A. (2009a). Heart sounds quality analysis for automatic cardiac biometry applications, *Proceedings of the 1st IEEE International Workshop on Information Forensics and Security*.

Beritelli, F. & Spadaccini, A. (2009b). Human Identity Verification based on Mel Frequency Analysis of Digital Heart Sounds, *Proceedings of the 16th International Conference on Digital Signal Processing*.

Beritelli, F. & Spadaccini, A. (2010a). An improved biometric identification system based on heart sounds and gaussian mixture models, *Proceedings of the 2010 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, IEEE, pp. 31–35.

Beritelli, F. & Spadaccini, A. (2010b). A statistical approach to biometric identity verification based on heart sounds, *Proceedings of the Fourth International Conference on Emerging Security Information, Systems and Technologies (SECURWARE2010)*, IEEE, pp. 93–96. URL: *http://dx.medra.org/10.1109/SECURWARE.2010.23*

Biel, L. & Pettersson, O. & Philipson, L. & Wide, P. (2001). ECG Analysis: A New Approach in Human Identification, *IEEE Transactions on Instrumentation and Measurement* 50(3): 808–812.

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

Davis, S. & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Transactions on Acoustics, Speech and Signal Processing* 28(4): 357–366.

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

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

Gravier, G. (2003). SPro: speech signal processing toolkit. URL: *http://gforge.inria.fr/projects/spro*

Jain, A. K., Flynn, P. & Ross, A. A. (2008). *Handbook of Biometrics*, Springer.

Jain, A. K., Ross, A. A. & Pankanti, S. (2006). Biometrics: A tool for information security, *IEEE Transactions on Information Forensics and Security* 1(2): 125–143.

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

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

McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions*, Wiley.

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

Prabhakar, S., Pankanti, S. & Jain, A. K. (2003). Biometric recognition: Security & privacy concerns, *IEEE Security and Privacy Magazine* 1(2): 33–42.

Rabiner, L., Schafer, R. & Rader, C. (1969). The chirp z-transform algorithm, *Audio and Electroacoustics, IEEE Transactions on* 17(2): 86 – 92.

Reynolds, D. A., Quatieri, T. F. & Dunn, R. B. (2000). Speaker verification using adapted gaussian mixture models, *Digital Signal Processing*, p. 2000.

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

Sabarimalai Manikandan, M. & Soman, K. (2010). Robust heart sound activity detection in noisy environments, *Electronics Letters* 46(16): 1100 –1102.

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