TÉCNICO LISBOA

Cancelable Biometrics

A dissimilarity based approach using hermite polynomial approximations

Afonso Alexandre Vicente Clara

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisor(s): Professora Ana Luísa Nobre Fred
Doutor Rui Cruz Ferreira

Examination Committee

Chairperson: Professor João Pedro Estrela Rodrigues Conde
Supervisor: Professora Ana Luísa Nobre Fred
Member of the Committee: Professor Paulo Luís Serras Lobato Correia

November 2015

To my family.

Acknowledgments

Em primeiro lugar, à professora Ana Fred, por todo o apoio que me prestou, por me ter guiado ao longo dos últimos meses a tomar as melhores decisões para o desenvolvimento deste trabalho. Ao Dr. Rui Ferreira e e ao Rui César das Neves por disponibilizarem os registos usados para realizar os testes. Ao Carlos Carreiras, por me ter esclarecido inúmeras dúvidas, permitindo-me desenvolver o trabalho a que me propus. À Diana Baptista, por me ter disponibilizado informação crucial ao trabalho.

Em segundo lugar, aos meus pais e avós por terem sido o meu pilar durante os últimos cinco anos, por sempre demonstrarem confiança em mim e por me acarinharem diariamente. Especialmente à minha irmã que suportou o sacrifício de lidar comigo todos os dias nos últimos anos. Também à Maria, por sempre demonstrar o seu afecto para comigo, essencial para me fazer seguir em frente.

Em último lugar, aos meus companheiros de biomédica por terem partilhado comigo muitos momentos, não só ligados ao curso mas também em outros contextos e que originaram tantas e tantas boas memórias.

A todos, o meu muito obrigado.

Resumo

Hoje em dia, a procura por sistemas de reconhecimento robustos e seguros é cada vez maior. Os sistemas biométricos canceláveis surgem como uma das possíveis respostas neste contexto. O presente trabalho propõe o desenvolvimento de um novo sistema deste tipo aplicado ao electrocardiograma, baseando-se em dois grandes ramos: a) a aproximação por polinómios hermiteanos; b) classificador baseado em dissemelhanças. A aproximação por polinómios hermiteanos é aplicada directamente ao electrocardiograma, permitindo obter uma nova representação, evitando armazenar a informação original. O classificador baseado em dissemelhanças pode ser aplicado a qualquer característica que cubra os critérios definidos pela biometria e realiza o processo de reconhecimento (autenticação e identificação) num espaço diferente do original, conferindo também segurança e robustez ao sistema desenvolvido.

O sistema proposto foi aplicado a um base de dados contendo registos correspondentes a 612 sujeitos, avaliando a influência de vários parâmetros relativos tanto à aproximação por polinómios hermiteanos, como ao classificador baseado em dissemelhanças. Esta metodologia foi comparada com um sistema clássico de biometria aplicado ao electrocardiograma, e com um sistema biométrico cancelável, também aplicado ao electrocardiograma.

O sistema apresentou um bom desempenho a nível da autenticação. Quando usadas as aproximações por polinómios hermiteanos verificou-se uma ligeira quebra no seu desempenho quando comparado com o uso dos registos originais. Usando uma técnica de *biometric menagerie*, foi possível tirar conclusões quanto à qualidade dos registos da base de dados usada.

Palavras-chave: Electrocardiograma, sistema biométrico cancelável, representação baseada em dissemelhanças, aproximação por polinómios hermiteanos.

Abstract

Nowadays, the demand for robust and secure identity recognition systems is increasing. Cancelable biometric systems emerge as one of the possible answers in this context. The present work aims to develop a new system of this kind applied to electrocardiogram (ECG) signals, based on two major techniques: a) exploring hermite polynomial approximations of the ECG; b) a dissimilarity based representation. The hermite polynomial approximation is applied directly to the ECG records, leading to a new representation, avoiding original data storage. The dissimilarity based representation can be applied to any characteristic which fulfils the criteria established by biometrics and its recognition process (authentication and identification) is done in a different space from the original one, granting security and robustness to the proposed system.

This proposed methodology was applied to a database with records regarding 612 subjects, evaluating the influence of several parameters related with both the hermite polynomial approximation and the dissimilarity based representation. The system was compared with a classic biometric system as well as with a cancelable one using electrocardiogram.

The system had a good performance regarding authentication. When the hermite polynomial approximations were presented as input to the system, a slightly decrease on performance was verified when compared with the system using the original records. Using a biometric menagerie technique, it was possible to make conclusions regarding the quality of the records contained on the used database.

Keywords: Electrocardiogram, cancelable biometrics system, dissimilarity based representation, hermite polynomial approximation.

Contents

List of Tables

List of Figures

List of Acronyms

ATM	Automated Teller Machine
PIN	Personal Identification Number
ECG	Electrocardiogram Signal
TP	True Positive
FN	False Negative
FP	False Positive
TN	True Negative
FAR	False Acceptance Rate
FRR	False Rejection Rate
ROC	Receiver Operating Curve
EER	Equal Error Rate
t_{ID}	Correct Identification
f_{ID}	Incorrect Identification
r_{ID}	Rejected Identification
EID	Identification Error
Pe_{ID}	Error Probability
RR_{ID}	Rejection Rate
F_{ID}	Total Number of Incorrect Identifications
T_{ID}	Total Number of Correct Identifications
R_{ID}	Total Number of Rejected Identifications
PC_{ID}	Identification Point of Compromise
k-NN	k Nearest Neighbour Classifier

SVM Support Vector Machine

FIR Finite Impulse Response

NRMSE Normalized Root-Mean Square Error

HSM Hospital de Santa Marta

Chapter 1

Introduction

1.1 Motivation

Nowadays, there is an increasing demand to have robust and secure identification/authentication systems, hereafter referred globally as recognition systems. This happens because there is an increasing number of processes which are not directly controlled by human beings. Several examples can be found: having access to our banking account through an automated teller machine (ATM) or internet; having access to a certain restricted area; or a device being able to adapt itself according to a given user. During a long period of time, these systems were based on credentials, such as the use of identification documents or personal identification numbers (PIN's). Obviously, these kind of procedures were not able to meet the growing demands regarding security in many applications, like border crossing, access control, financial transactions or telecommunications [14]. In fact, it is estimated that credit card fraud has values around $450 million per annum in the United States of America and a part of this value is due to stolen or lost credit cards, and about 1 billion dollars is the value resulting from telephone calls made by bandwidth thieves [14]. In 2012, according to the European Central Bank [1], €794 million was the value regarding frauds resulted from card-not-present payments. Another example is the number of illegal immigrants that cross the border between Mexico and the United States of America with stolen documentation [14].

Thus, it is clear that other forms of recognition systems needed to be found in order to provide answers related to the problems mentioned above. One of the fields of study which emerged as a possible answer was the biometric systems field. The main purpose of these kind of systems is to use a certain characteristic intrinsic to the human being in a context of a recognition process. A more formal definition is given in the next chapter. This approach presents some advantages. For example, biometric information cannot be lost or forgotten in contrast to cards or PIN's. It is not easy to forge this kind of information and from the user's perspective, it is not necessary to remember any password. On the other hand, when the system is compromised and someone gains access to the database information, there is a permanent biometric data compromise, since this kind of information cannot be revoked [27].

To give an answer to the revocable issues, cancelable biometric systems have emerged. This ap-

proach has two main purposes: on the one hand, tries to develop recognition systems more difficult to forge or copy; on the other hand, it is able to protect the user's information since that information is not directly stored on the database, granting confidence to the system. From this perspective, it is possible to point some advantages of these systems. They are more secure and more difficult to crack than the traditional biometric systems. Besides that, it could be expected that these kind of systems raise more social acceptance in relation to the traditional biometric systems, since the biometric data is not kept in the database, proving privacy to the users.

1.2 Proposed Methodology and Thesis Goal

The main purpose of the present work is to develop a new cancelable biometric system able to give a response to the problems raised above. Thus, two main fields will be explored. On the one hand, a dissimilarity based representation is used in the process of recognition. On the other hand, transformations based on hermite polynomial approximation are applied to the electrocardiogram (ECG) waveform, in order to modelling QRS complex. These two techniques are able, separately or set together, to guarantee original data protection. It is important to note that the dissimilarity based approach can be applied not only to the ECG signal, but to any other signal which may be used to perform subject recognition.

Generically, the applied methodology can be described as follows:

- ECG signal pre-processing, involve filtering and signal segmentation into heartbeats.
- Hermite polynomial approximation is applied to the segmented signals.
- A decision regarding recognition is made, by applying a classifier based on a dissimilarity representation.

Besides that, in order to evaluate the database's inter-variability and intra-variability, a biometric menagerie technique is applied to the proposed method.

1.3 Thesis Contributions

As a consequence of applying the methodology described above, a set of contributions were made. Regarding hermite polynomial approximation, the influence of the number of hermite polynomials on the systems performance was studied. Regarding dissimilarity based approach, the influence of the measures used was studied by first comparing genuine/impostor distributions and next by applying a classifier to the database used, in order to observe the influence of the population percentage used to built the training dataset.

The proposed approach was applied to a data set in order to evaluate its performance. Besides that, these results were compared with other methodologies reported in the literature in order to evaluate its effectiveness.

By applying the biometric menagerie technique it was possible to observe which subjects belonging to the used database had characteristics that allowed a good and a bad performance at the recognition moment.

1.4 Thesis Structure

This thesis has 6 chapters organized as follows: Chapter 1 presents the thesis' motivation, presents its goals and methodology proposed to give an answer to the explained problem, and finally refers its contributions; Chapter 2 introduces the essential concepts mentioned through the present work, essential to its understanding; Chapter 3 approaches the state of the art regarding biometric systems and cancelable biometrics, presenting an overview of the main works developed in this area of investigation; Chapter 4 presents the proposed methodology as well as some concepts regarding it; Chapter 5 presents the results of applying the proposed method in comparison with other methods present in the literature; finally, in Chapter 6 an overall conclusion is done, presenting ideas for future works in the area.

Chapter 2

Founding Concepts

2.1 Biometric Systems

As said in section 1.1, the biometric systems' field aims to develop recognition systems based on human characteristics. More formally, one can define biometrics as automated recognition of individuals based on their behavioural and/or physiological characteristics (ISO/IEC JTC1 SC37).

Not all human characteristics (behavioural/physiological) are eligible to be considered as a biometric feature. To be applied to a biometric system, the behavioural or biological characteristic has to fulfil some properties, namely [27]:

- *Universality* - every subject has that characteristic.
- *Uniqueness* - there are not two persons with the same terms of a given characteristic.
- *Permanence* - the characteristic is invariant along time.
- *Collectability* - the characteristic can be measured quantitatively.
- *Performance* - the ability to, given a certain characteristic, the system performs accurately.

In practice, there are other criteria which are associated with the specificity of a certain characteristic, namely: acceptability or circumvention, all of them related with how good is a certain characteristic to be used to distinguish an individual [4]. Examples of human characteristics which can be applied to a biometric system are: fingerprint, palm print, signature, face, iris, voice or hand geometry.

2.1.1 System's Architecture

The architecture of a biometric system follows a typical design of a pattern recognition system. Thus, its basic modules are: signal acquisition; pre-processing, which can include de-noising and other pre-processing tools and feature extraction, for example through a transformation.

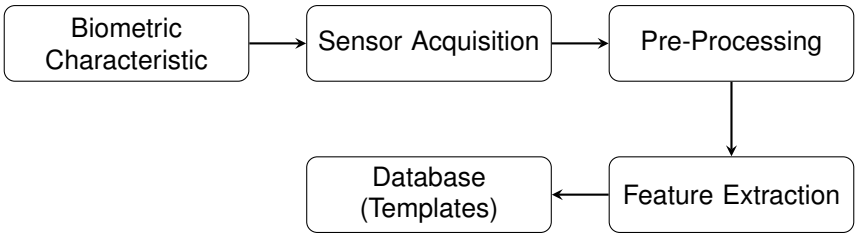

Figure 2.1: Biometric system architecture, when working on enrolment mode.

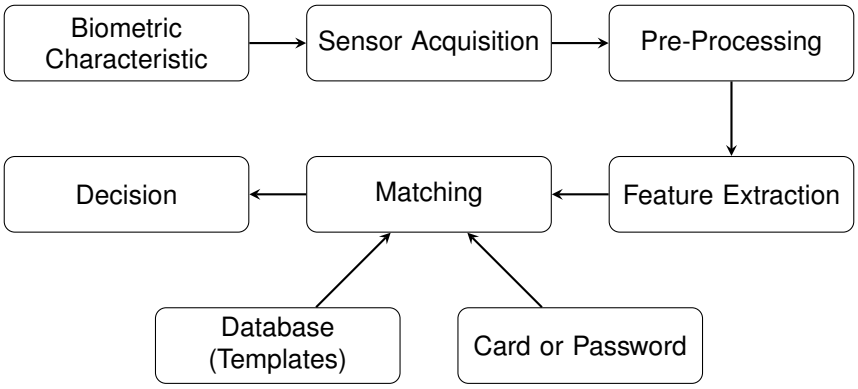

Figure 2.2: Biometric system architecture, when working on authentication mode.

2.1.2 Enrolment and Recognition Phases

A typical biometric system has two different operation modes or phases: enrolment and recognition which can be an authentication or an identification process.

At the enrolment phase (see figure 2.1), the system has the ability to build a database from data belonging to a certain subject to whom identity is known and also provided to the system. For each subject, a certain number of templates can be generated as a way to best represent that individual [13].

In authentication mode (see figure 2.2), a subject provides to the system his/hers biometric measurement and a claimed identification, and the system tries to match that data only with the stored one referring to that subject, and come up with a decision [13]. This decision is based in a threshold, meaning if the resulting distance measure regarding the comparison between the provided data and the template(s) is below of a certain defined threshold, the subject is accepted by the system.

From the comparison between the stored templates and the test data, two scores can be generate, namely genuine and impostor scores. On the one hand, genuine score, or distribution, is obtained by taking the measures between the template(s) and the test data from the same subject, for all subjects. On the other hand, impostor score is accomplished by taking the measures between the template(s) from a certain subject and the test data regarding all the other subjects.

In identification mode (see figure 2.3), the system measures a certain biometric characteristic from the subject, tries to match the data with the one stored and eventually is able to identify the subject [13]. This operation mode is more susceptible to errors, since the system needs to compare the provided data whit the data belonging to all subjects from the database.

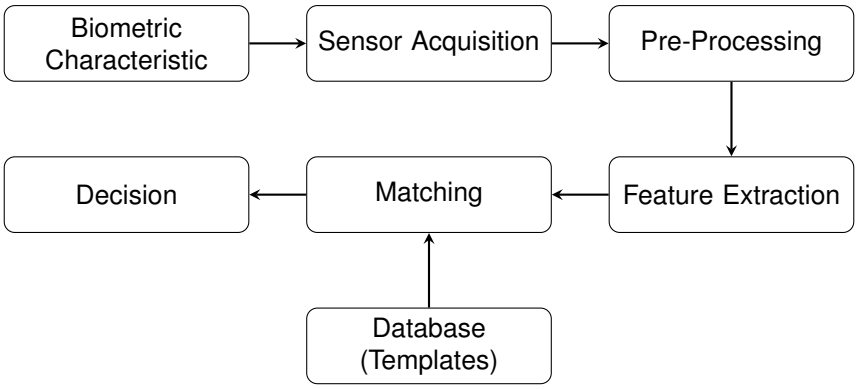

Figure 2.3: Biometric system architecture, when working on identification mode.

2.2 Performance

When one develops a biometric or cancelable biometric system, it is important to evaluate its performance. Its accuracy depends not only on the system's design but also on the data used to test the system [13]. Thus, when one is designing this kind of systems, it is important to keep in mind that there is not a perfect configuration and the design should reflect the environment for which the system is being developed. In order to test the system's performance, at the authentication time, it is possible to compute some parameters. When a system is being tested, four possible scenarios can occur [13]: a genuine individual is accepted, meaning a true positive occurred (TP); a genuine individual is rejected, a false negative (FN); an impostor individual is accepted, a false positive (FP); or an impostor individual is rejected, a true negative (TN). With these four measures, it is possible to calculate some rates, namely the false acceptance rate (FAR) and the false rejection rate (FRR), which are defined respectively by:

$$FAR = \frac{FP}{TN + FP} \tag{2.1}$$

$$FRR = \frac{FN}{TP + FN} \tag{2.2}$$

Thus, it is possible to conclude that smaller threshold values will favour smaller FAR values while larger threshold values will favour smaller FRR values, meaning that a compromise between these two measures should be obtained. As a way to provide information about the system's performance over different operating points, another measure is used, related with FAR and FRR, called receiver operating curve (ROC). These plots give another important quantity, the equal error rate (EER), defined as the value where FAR and FRR are equal, which can be interpreted as the compromise point between them (see figure 2.4).

In order to evaluate a system when working on identification mode, two different scenarios can be evaluated. On the one hand, this process can be done without threshold application. Thus, when the system tries to identify a subject, it searches for the most closely related template with that subject. Two possibilities can occur: the system correctly identifies the subject, generating a correct identification, t_{ID} or it fails to identify the subject originating a wrong identification, f_{ID}. On the other hand, the system can use a threshold. Therefore, when trying to identify a subject, the smallest distance between the provided

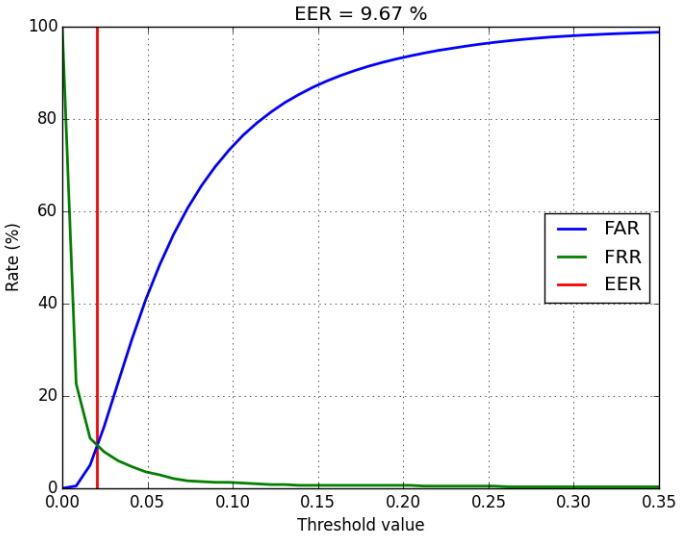

Figure 2.4: An example of the evolution of FAR and FRR as well as its intersection representing EER value.

data and the templates, d_i is compared to the threshold value, t. Three different situations can occur:

- $d_i < t$ and d_i is taken from a template belonging to the subject. Thus a t_{ID} is generated.
- $d_i < t$ and d_i is taken from a template not belonging to the subject. Thus a f_{ID} is generated.
- $d_i > t$, meaning the system does not attribute an identification to the subject, thus being rejected, r_{ID}.

In this way, it is possible to compute three different measures in order to evaluate the identification process: the identification error, EID, the error probability, Pe_{ID} and the rejection rate, RR_{ID}. In the case of the system does not consider any threshold, RR is not defined and the remaining two measures are defined in the same way, as follows:

$$Pe_{ID} = EID = \frac{F_{ID}}{T_{ID} + F_{ID}} \tag{2.3}$$

In the case of a threshold is being considered, then the three measures are defined respectively by:

$$EID = \frac{F_{ID}}{T_{ID} + F_{ID}} \tag{2.4}$$

$$Pe_{ID} = \frac{F_{ID}}{R_{ID} + T_{ID} + F_{ID}} \tag{2.5}$$

$$RR_{ID} = \frac{R_{ID}}{R_{ID} + T_{ID} + F_{ID}} \tag{2.6}$$

where F_{ID} is the number of incorrect identifications and T_{ID} is the number of correct identifications and R_{ID} is the number total rejections. In the same way as for authentication, there is also a compromise between Pe_{ID} and RR_{ID}, hereafter denoted as PC_{ID} (see figure 2.5).

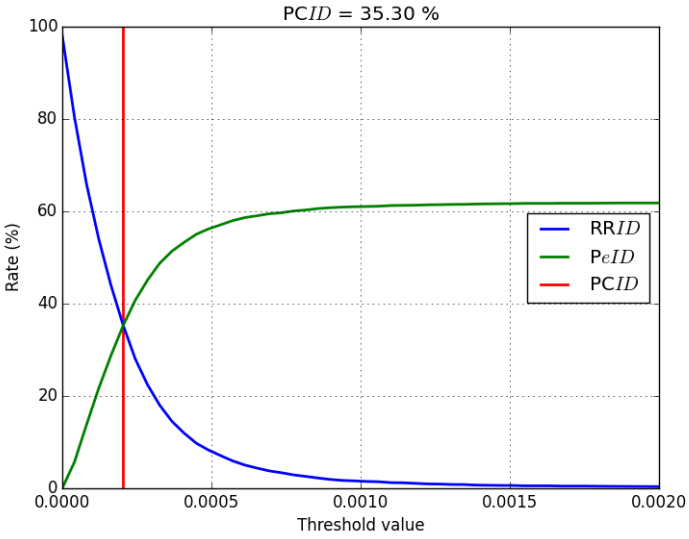

Figure 2.5: An example of the evolution of RR_{ID} and Pe_{ID} as well as its intersection representing PC_{ID} value.

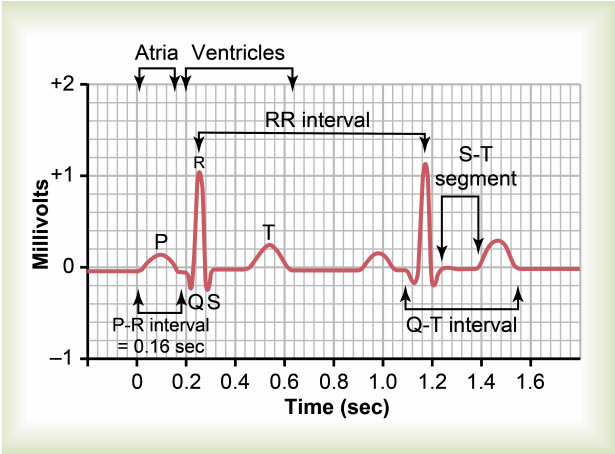

Figure 2.6: An ECG waveform. Current figure was extracted from [10].

2.3 Electrocardiogram

ECG is the recording of the electrical activity of the heart. Through ECG it is possible to observe the phenomena related with the polarization/depolarization of the cardiac cells. These phenomena are represented in ECG by waves/complexes, namely (see figure 2.6): **P wave**, which is caused by the atria depolarization; **QRS complex**, which is caused by the ventricle depolarization; **T wave**, which is caused by the ventricular repolarization [10].

One of the most common ways to measure heart's electrical activity is through the Einthoven's triangle (see figure 2.7). This triangle is built by placing the electrodes in both arms (usually on the wrists) and on the left leg (electrode placed on the ankle). Thus, it is possible to obtain the voltage between two of them, usually termed as *lead*. From the Einthoven's triangle, one can obtain three leads [10].

Another way to acquire ECG, probably the most used, is through the placement of electrodes on the chest. This method allows to obtain more leads, since usually one can place six electrodes on the chest,

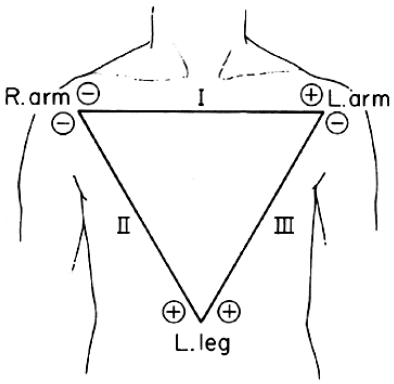

Figure 2.7: The Einthoven's triangle and its leads

near to the heart. These leads are known as precordial leads and are represented by V_1, V_2, V_3, V_4, V_5 and V_6 [10].

2.3.1 ECG as a Biometric

The use of ECG in biometrics is based on three assumptions [20], under the criteria mentioned above (see section 2.2):

- ECG is difficult to counterfeit in supervised conditions.
- It is present in all living individuals, during their entire life (*universality*).
- Easiness to collect (*collectability*).

As any other characteristic used to develop a biometric system, one of the challenges to overcome is related with the question of performance, since it is necessary to prove that a given biometric system based on ECG is able to exhibit good results. Besides the conditions mentioned above, ECG also provides additional information related to psychological states, and physiological and clinical status, which can be used as a complementary elements to the recognition process.

Using ECG as base to a biometric system was first proposed by Biel et.al [2]. Their experiments showed that it is possible to identify a person using just its ECG, since this signal has enough information to discriminate a subject from others.

As referred above, the ECG recordings are obtained by placing an electrodes set on a person's chest. Since the electrodes are placed on the skin surface, it is obvious that the voltages recorded will vary with several parameters, including skin impedance, the inhomogeneity of the thorax, the blood resistivity, the effect of respiration, or the kind of electrodes used in the measurements. Thus, one can expect that the ECG is highly individual and has sufficient quality to be used in a recognition system [21]. On the other hand, the heartbeat variability, can have a negative impact on the identification/authentication process, since it introduces intra-subject variability due to, for example, emotional states or the moment of day when the ECG is measured [23]. This topic is related with *permanence*. As seen, it is another challenge to answer in order to develop a real world application based on ECG.

Regarding the system's architecture, it is very similar to the modules described for a typical biometric system. More details related with each module will be given on the next chapter, including a concrete example.

2.4 Cancelable Biometrics

As mentioned in section 1.1, cancelable biometrics emerged as an answer to the issues to which traditional biometric systems were not able do deal with. One of those problems, is the revocability question, meaning, the biological or behavioural characteristic used by a given system usually cannot be changed if the database where it is kept are attacked or stolen [27]. Two main fields emerged as a possible solution: on the one hand, *biometric cryptosystems*. According to [26], these systems are designed to securely bind a digital key to a given biometric feature or generate a digital key from it. Based on how the key is generated or retrieved the used schemes are named as key-generation schemes or key-binding schemes, respectively; on the other hand, *cancelable biometrics*, which is the field the present work will explore. The two mentioned approaches are commonly referred in literature as **biometric template protection methods**.

The concept of cancelable biometrics was first introduced by Ratha et al. [24], as a way of applying intentional and repeatable distortions/transformations to a biometric signal to ensure the privacy of biometric data. Ideally, these distortions/transformations are non-invertible. In this way, one can store the transformed version of a biometric characteristic and hence providing higher privacy.

According to [26], there are two main ways to generate cancelable templates: through a **non-invertible transformation** or through **biometric salting**.

When using a non-invertible transformation, all the recognition process is performed, not in the original space, but in the space generated by the transformation. The main advantage of this approach is the difficulty to reconstruct the biometric data by an impostor, even when the transformation is compromised. On the other hand, this approach implies a loss of accuracy since there is frequently loss of information when applying the transformation. Biometric salting usually uses invertible transforms. Since the original information can be recovered by applying the inverse transformation, one must keep in secrecy the transform parameters. This can be done by giving to a certain user a key or password with the needed information. The main advantage of this approach is that it is able to maintain the recognition performance. On the other hand, it is less secure than the previous one [26, 15].

Cancelable biometrics templates are designed according to two major criteria [26, 27]:

1. *Irreversibility* - Meaning it should be easy to generate a cancelable template, through a **transformation or distortion**, but should be computationally hard or impossible to reconstruct the original data, hence protecting it.

2. *Unlinkability* - Meaning that system has the ability to generate different cancelable templates from the original data, ensuring **renewability** in case of an attack. Additionally, by crossing several cancelable templates originated by the same data, one cannot have information about that data.

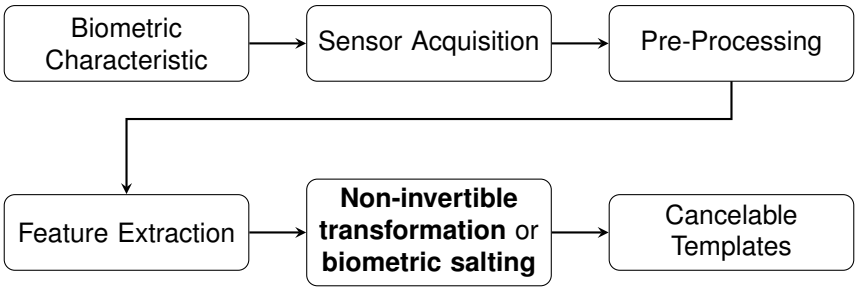

Figure 2.8: A typical cancelable biometric system. Note to the module which performs a Non-invertible transform or biometric salting.

The design of this kind of recognition systems is very similar to the one described above. Additionally, there is a module which is responsible by applying a non-invertible transformation or to perform salting (see figure 2.8).

2.5 Biometric Menagerie

As said above, it is usual to compute some parameters, like EER or PC_{ID} in order to evaluate the system's performance. These parameters are directly affected by, not only the proposed system, but also by the subjects of the tested population. Thus, it is useful to have an overview of the subjects' behaviour under the system's conditions, since there may be subjects with difficulties at the authentication time or subjects vulnerable to impersonation.

Since it is important to guarantee that the subjects belonging to a database have a consistent performance under the classification process, biometric menagerie aims to observe the influence of each subject in the classification process by dividing them into several categories according to its behaviour under the system.

This technique consists in assigning a certain animal-like behaviour to the subjects. According to [29], a certain subject can belong to one of five different animal groups. These groups are defined in terms of a relationship between genuine and impostor match scores. In order to define these scores, a few measures can be computed. According to [29], any well-defined statistical measure of performance can be used and its use is depending on the type of the biometric system one is using. All these quantities have advantages and drawbacks. The present work uses a simple approach, with the system working on identification mode. On the one hand, the genuine performance is based on the count of how many user's sample tests are indeed from an authorized user, meaning the samples which are correctly classified. On the other hand, the impostor performance is based on the number of sample tests regarding an user that the system considers to be forged or unauthorized, meaning the samples which are incorrectly classified. The different animal groups are [29]:

1. **Sheep** - Theoretically, most of the population belongs to this group, which is characterized by middle values of impostor and genuine scores. Thus, on average, these subjects will tend to match well against themselves but poorly against others.

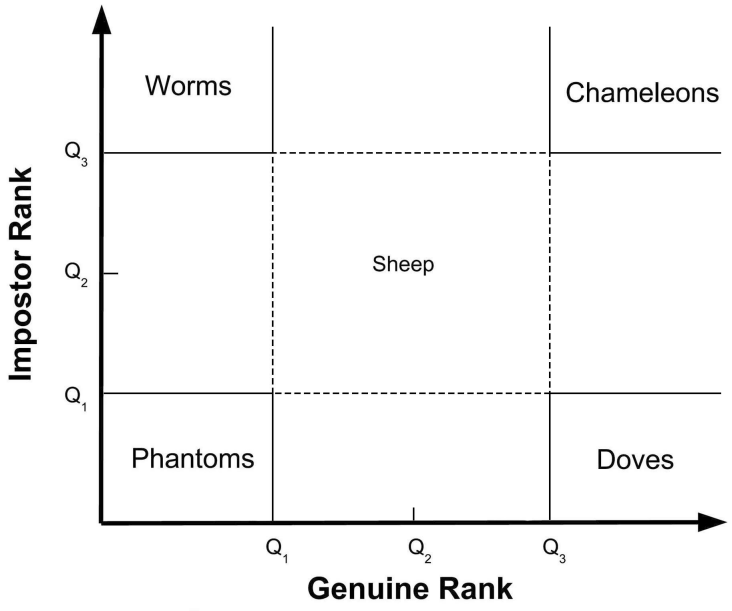

Figure 2.9: The behaviour of the several animal groups. The boarders are set by the quartiles Q_1 and Q_3. Current figure was extracted from [29].

2. **Worms** - These individuals are characterized by an high impostor score and a low genuine score, meaning they will tend to match against other subjects easily and perform badly when matched against themselves. Thus, they will contribute to decrease the classification quality.

3. **Phantoms** - This group has both low genuine and impostor scores. Thus, it is hard to match them against themselves.

4. **Chameleons** - This group is characterized by both high genuine and impostor scores, meaning that they easily match against other subjects, besides the good behaviour when they match against themselves.

5. **Doves** - In an ideal scenario, all the subjects should belong to this animal group since it presents low impostor scores but high genuine scores. Therefore, these subjects perform very well when matched against themselves and it is hard to match them against other subjects.

In the scope of the present work, the assignment of a certain subject to each of the animal groups was made in two simple steps:

1. Plot the number of genuine scores vs. impostor scores relative to all the subjects. The scores relative to each subject were normalized by the number of test samples regarding that user.

2. Quartiles were computed according to the statistical distribution. Quartiles Q_1 and Q_3 define the boarders of each animal group, in both dimensions (see figure 2.9).

2.6 Hermite Polynomials

One of the main drawbacks of ECG is the big amount of information that records produce. It is not unusual to record a person's ECG during long periods of time, for example during his/her sleeping. Many times, it is also necessary to inspect records with a great detail, which is a very hard task to do.

The use of hermite functions to represent heartbeats emerged as a new method to provide a very compact representation of ECG, namely by using just the coefficients resulting from the linear combination of the hermite polynomials which approximate the QRS complex as features in an automatic classification system.

The present work aims to look not just to the coefficients of the linear combination of the hermite polynomials but to the entire hermite polynomial approximation.

According to [18], hermite polynomials will provide a better characterization of the heartbeat if the point of maximum symmetry is selected as the center of the window of signal to be fitted. In a normal heartbeat this point is the peak of the the QRS complex, which corresponds to the R wave. Thus, in the present work, hermite polynomials will approximate the QRS complex from a heartbeat.

A QRS complex, $x(t)$, which to be approximated by the hermite polynomials, can be described as:

$$x(t) = \sum_{n=0}^{N-1} c_n(\sigma)\phi_n(t,\sigma) + e(t) \tag{2.7}$$

where N is the number of hermite polynomials used in the approximation, $\phi_n(t,\sigma)$ is the n^{th} hermite function, $c_n(\sigma)$ are the coefficients or weights of the linear combination associated with each function, σ is the parameter which controls the width of the functions and $e(t)$ is the error associated with the approximation.

The n^{th} hermite function is defined as (see figure 2.10):

$$\phi_n(t,\sigma) = \frac{1}{\sqrt{\sigma 2^n n! \sqrt{\pi}}} e^{-t^2/2\sigma^2} H_n(t/\sigma) \tag{2.8}$$

$H_n(t/\sigma)$ is the n^{th} hermite polynomial. Hermite polynomials are recursively given by:

$$H_0(x) = 1 \tag{2.9}$$

$$H_1(x) = 2x \tag{2.10}$$

$$H_n(x) = 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x) \tag{2.11}$$

The values of the width parameter σ for 3, 4, 5 and 6 polynomials are showed in [16] and can be seen in figure 2.1. From 6 polynomials, the value of σ is constant and equal to 47ms.

Determine the coefficients $c_n(\sigma)$ is a problem of minimizing the summed square error

$$\sum_t |e(t,\sigma)|^2 = \sum_t |x_t - \sum_n c_n(\sigma)\phi_n(t,\sigma)|^2 \tag{2.12}$$

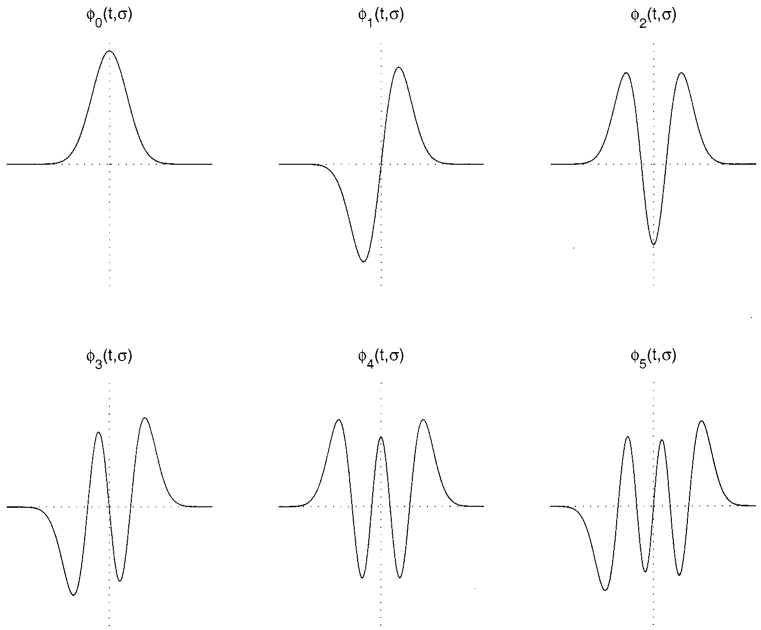

Figure 2.10: The first six hermite functions. Current figure was extracted from [16]

Table 2.1: Values of σ for the number of hermite polynomials used to model the QRS complex.

Hermite Polynomials	σ **(ms)**
3	62
4	55
5	51
6	47

2.7 Dissimilarity Representations and Classifiers

In order to make a decision about a given subject, all the biometric systems use a classifier. Since any classifier has advantages and drawbacks, one can chose several different classifiers to be part of the system such k nearest neighbour classifiers (k-NN's), support vector machines (SVM's) or neural networks.

Typically, classifiers use as base data representation a feature space. In this thesis, inspired by the work of [7, 9], it is proposed the use of a dissimilarity based representation, to which the classifiers are built upon.

As the name suggests, the main purpose of this classifier is to perform a decision based on dissimilarities between objects. As suggested by [8], generally human brain is first triggered by the differences between objects and just after that comes a feature based description if a certain subject is trying to name a given object. This fact serves as motivation to explore this new type of data representation. Hereafter, the classifier using the dissimilarity representation as input will be referred as dissimilarity based classifier.

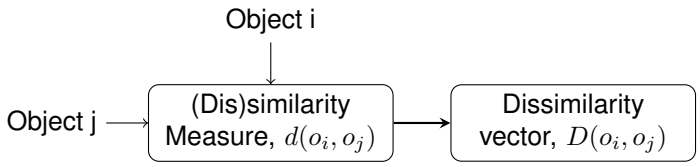

Figure 2.11: Dissimilarity vector generation.

2.7.1 Dissimilarities

Before describing the system, it is important to define the dissimilarity measure concept and some of its properties. According to [7], a dissimilarity measure $d(o_i, o_j)$ between two objects o_i and o_j can be seen as the degree of difference between them and has several properties, such as:

- *Non-negativity*: $d(o_i, o_j) \geq 0$;
- *Identity of indiscernibles*: $d(o_i, o_j) = 0$ if and only if $o_i \equiv o_j$;
- *Symmetry*: $d(o_i, o_j) = d(o_j, o_i)$.

Thus, one way to extract dissimilarities is by using a similarity measure. There are several similarity measures which fit into the criteria defined by [7]. In the scope of the present work, the measures that will be used are the **euclidean distance** and the **cosine similarity**. This choice is based on two main facts: on the one hand, the ECG shape suggests that the cosine similarity is a good shape similarity measure to apply at the extraction moment; on the other hand, one of the most used measures in k-NN's is the euclidean distance. If two objects are represented by two vectors o_i and o_j respectively, the euclidean distance and cosine similarity can be defined as

$$D(o_i, o_j) = \sqrt{(o_i - o_j)^T (o_i - o_j)} \tag{2.13}$$

$$D(o_i, o_j) = \frac{o_i . o_j}{\|o_i\| \, \|o_j\|} \tag{2.14}$$

2.7.2 Dissimilarity Representation

As said above, this classifier suggests performing a decision in a different space from the original one. The approach followed to extract dissimilarities is explained in [9] and it is named **inter-subject approach**, meaning that dissimilarity based representation assumes the existence of n representative objects (in the limit, all the training data), which can be called *prototypes*. Prototypes are randomly selected from the training data, assuming that they can cover all the population variation. Thus, the representation of a single object is given by the vector of dissimilarity of this object to each of the prototypes. Finally, it is possible to define the dissimilarity space, D_S resulting from all the vectors of dissimilarity of all the objects to each of the prototypes.

In the present work, the objects are considered to be the ECG heartbeats, in the case of applying the system solely to records, and the hermite polynomial approximations of the QRS complexes.

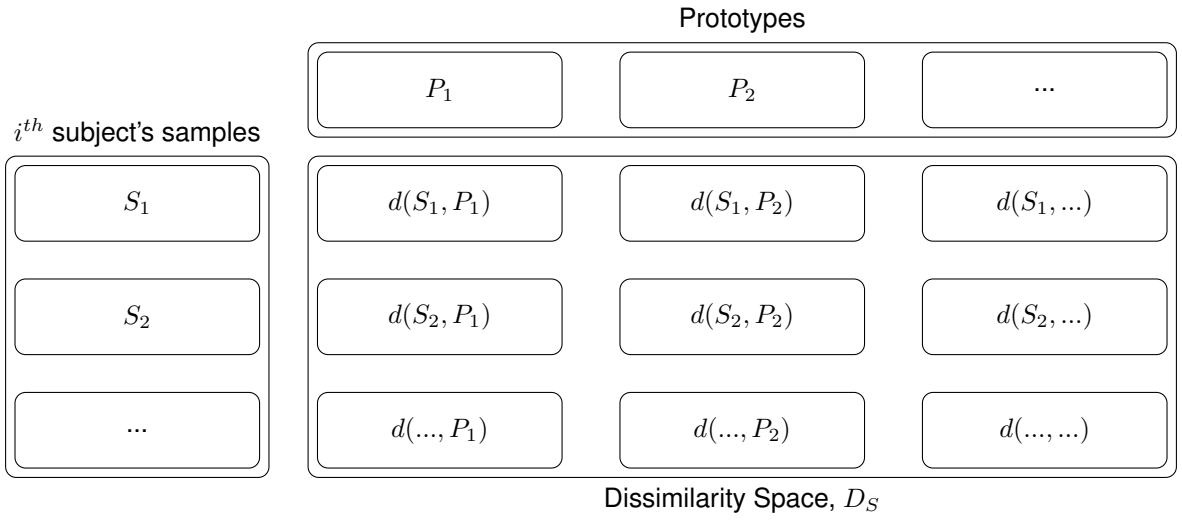

Figure 2.12: Dissimilarity space generation.

Chapter 3

State of the Art

3.1 ECG Biometric Systems

As mentioned in section 2.3.1, an architecture for a biometric system using ECG (see figure 3.1) follows, generally speaking, a typical architecture for a pattern recognition system. It is important to keep in mind some of its modules and their main goals. Therefore, more detailed information will be given regarding it.

The majority of the ECG biometric studies are based on one-channel ECG [21]. These recordings are, often, contaminated with noise, both as in high as in low frequencies. The low frequency components are usually associated with changes in baseline electrical potential of the device, while the high frequency components are associated with electrical/magnetic field of building power and the process of digitization of the signal [12]. To remove the noise, it is applied a band pass filter, usually, a finite impulse response (FIR) filter with cut-off frequencies between 5Hz and 20Hz [3].

Besides de-noising, the pre-processing stage includes, often, ECG segmentation into single heartbeats. One of the reasons to do it is that the variations across individuals within one cardiac cycle is thought to be sufficient in discriminating amongst them [21].

ECG biometric methods can be divided in three classes: fiducial, non-fiducial and hybrid or partially fiducial methods. Fiducial methods are based on extracting features from characteristic points within an heartbeat, such as wave amplitudes, slope information or time between two waves. On the other hand, non-fiducial methods do not use characteristic points to extract features. Instead, feature extraction is done on another space, for example in the frequency domain. Finally, hybrid methods combine both fiducial and non-fiducial methods [21, 5].

The work of Coutinho et.al [5] provides an example of a fiducial approach as well as an example of a non-fiducial approach. They collected approximately 6min of ECG data for each subject. During the data acquisition, the subjects were asked to complete a task requiring mental concentration.

Regarding the fiducial approach, they obtained a mean waveform for groups of ten consecutive heartbeat waveforms. From each of the mean wavesforms, they were able to extract four amplitude features, namely the amplitudes of P, Q, S and T waves and four time features also associated to the

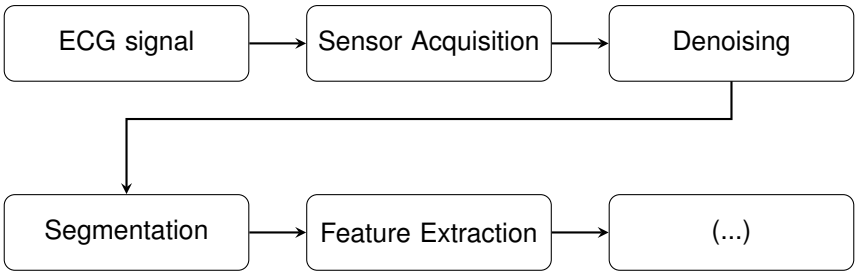

Figure 3.1: Part of biometric system based on ECG. The last module follows the architecture depicted in Figure 2.1

waves referred. Besides that, they subsampled each one of the mean waverforms in order to obtain 45 more amplitude features. Those features served as input to the used classifier, which was a k-NN with k set to 1, using the euclidean distance as a decision criteria.

In relation with the non-fiducial approach, they proposed the conversion of ECG signals into a string, by using quantisation adapted to each subject through the Lloyd-Max algorithm. After the quantisation process, the Lempel-Ziv algorithm was used as a string matching tool. Roughly speaking, in order to classify a given sample, the algorithm would generate a description of all the database templates based on the given input string. Therefore, the input string would be classified as belonging to the subject related with the shortest description.

The work of Wang et.al [28] is an example of an hybrid approach. They proposed an hierarchical scheme. First, a Wilks' Lambda-based stepwise method was used to extract 15 temporal features and the classification was done by using linear discriminant analysis. If all the heartbeats were classified as belonging to the same subject, the process would stop in this step. If not, a principal component analysis based classification would be applied to the subjects whom were wrongly classified.

Lastly, the classifiers used are prevailingly k-NN, SVM or neural networks but others can be used, such as statistical based approaches [21].

3.1.1 A Specific Example

As a way to provide a concrete example of a biometric system using ECG, the work of Carreiras et.al [3] is referenced here.

Their purpose was to study the *uniqueness* question regarding ECG. The proposed biometric system includes the typical architecture described on the section above and can be summarized in the following way (see figure 3.2):

1. ECG acquisition from one of the ECG leads (in this case, lead I).

2. The raw data passes through a preprocessing block which includes a FIR filter (order 150) with cutoff frequencies between $[5; 20]$Hz, and QRS complex detection.

3. Next, an algorithm for outlier detection is applied. This algorithm is called DMEAN. After computing the mean template regarding a record, the algorithm decides if an heartbeat is an outlier if its distance to the mean template is greater than a certain threshold.

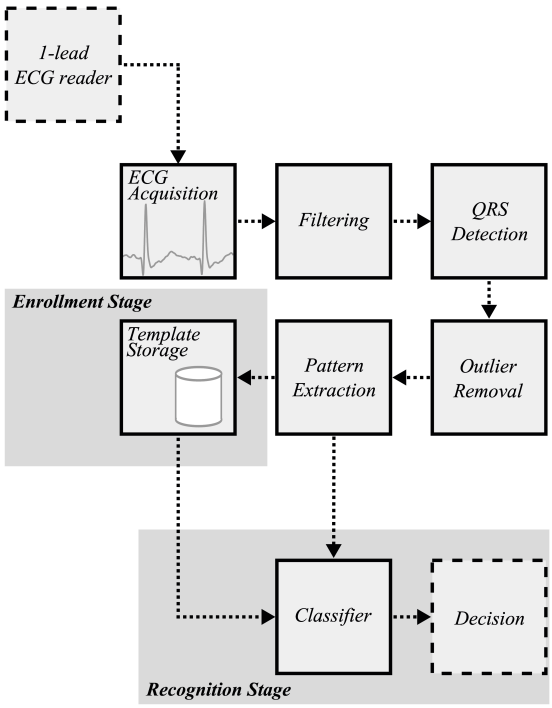

Figure 3.2: Biometric system's architecture applied by Carreiras et.al [3]. Current figure was extracted from [3].

4. At the pattern extraction block, all the heartbeats are aligned by their R-peak instants and a 600ms window is applied to them. All the amplitudes in this interval will be used to feed the classifier.

5. Finally, a k-NN is applied with k set to 3 and cosine similarity used as classification metric. Three heartbeats from each subject are used as templates and the testing phase is performed using one heartbeat from each subject.

3.2 Cancelable Biometrics

In spite of being a recent topic, there are already some interesting studies in this field of study. These approaches are focused mainly on the fingerprint, iris and handwriting/signature.

Nagar et.al [19] proposed a method to generate non-invertible fingerprint templates and a new measure of non-invertibility which they called coverage-effort, able to compute a degree of how a template is non-invertible. In order to generate cancelable templates they followed the strategy proposed by Ratha et.al [25] based on an one-way transformation in feature domain. As in [25], they applied three different transformations, namely: cartesian transformation, polar transformation and surface folding transformation (see figure 3.3). All of them were applied to the minutiae which are the points that best give a distinctive representation of fingerprint since these are the points where the friction ridges end or bifurcate [19]. To compute the coverage-effort curve, three steps were taken [25]:

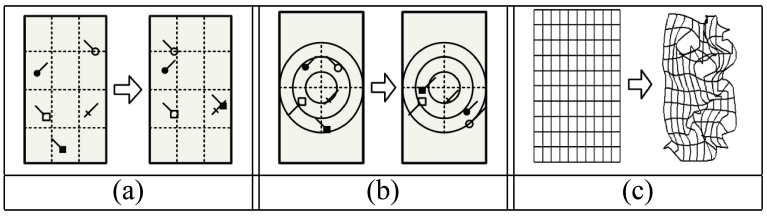

Figure 3.3: Feature transformation applied by [19]. (a) represents the Cartesian transform. (b) represents the polar transform. (c) represents the Gaussian mixture transform. Current figure was extracted from [19].

1. *Pre-image Computation* - compute the pre-images transformed minutiae in order to transformation lead to the given transformed minutia.

2. *Minutiae Likelihood Computation* - estimation of the relative probability of a certain minutia in the pre-image using a kernel density estimation.

3. *Non-invertibility Measure Computation* - by sorting the pre-images according to their likelihoods, it is possible to compute the coverage, meaning, the number of true pre-images guesses.

By computing this parameter, it is possible to chose between the several non-invertible templates, those which present the lowest risk security and the better matching performance.

Ouda et.al [22] proposed a new cancelable biometrics method regarding iris. According to them, this method satisfies all the requirements of a cancelable biometrics system, namely *revocability* and *noninvertibility*, maintaining the recognition performance when compared against a traditional biometric system. Their approach consists in three main steps (see figure 3.4): IrisCode generation from iris' images; extraction of the consistent bits from the IrisCode; and cancelable code generation. They considered a bit to be consistent if it does not change across the several irisCodes extracted from the iris' images. To generate a cancelable code, a given IrisCode is divided into bit words of a fixed length. These words are mapped into a new vector. Its position in the vector is defined by a random sequence generated from a specific seed associated to each user. This vector was named BioCode and it is stored as a cancelable template. They also studied the process's noninvertibility. They proved that it would be necessary a number of trials around 2^n if an attacker wanted to have access to the original IrisCodes, with n being the number of bits of a given IrisCode. Thus, they developed a method with a considerable security degree.

Maiorana et.al [17], proposed a new scheme named BioConvolving, with the capability to be applied, according to them, to any biometrics. They applied this method to an on-line signature recognition system. The basic idea of their system was to randomly split the original signal into non-overlapping parts, according to a predefined key and to perform next a linear convolution between those signal parts (see figure 3.5). To introduce a higher degree of renewability, they proposed two approaches. On the one hand, besides the key which is used to split the signal, they proposed an additional transformation key which is responsible to randomly select different parts from different signals. Therefore it is possible to convolve parts from different signals. On the other hand, a shift parameter can be randomly generated.

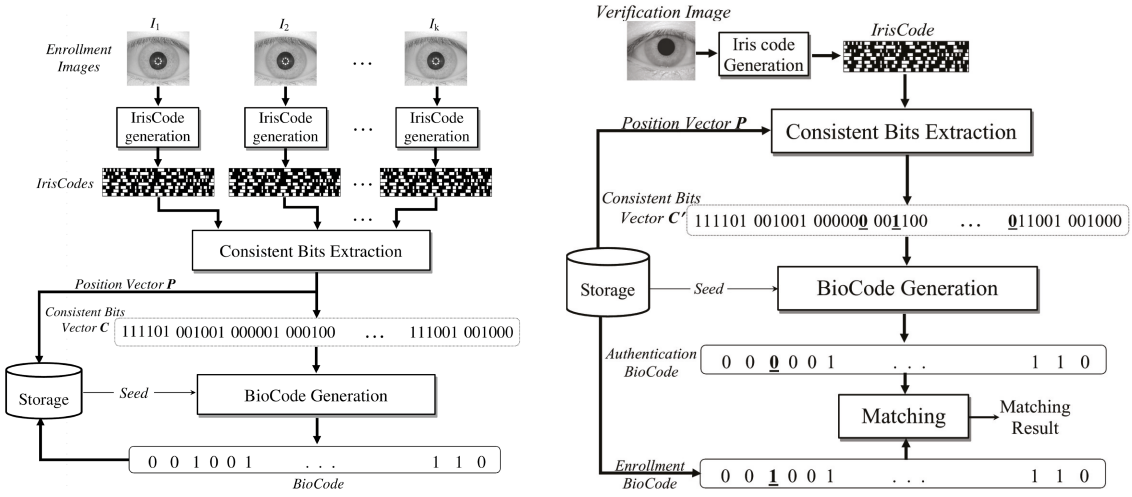

Figure 3.4: Proposed scheme by [22]. On the left, the enrolment phase. On the right, the verification phase. Current figure was extracted from [22].

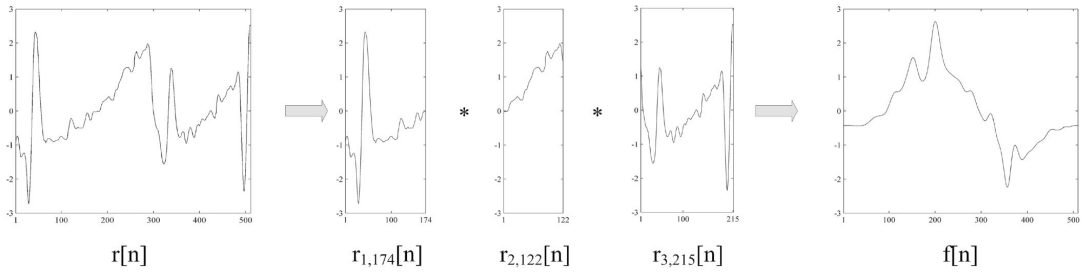

Figure 3.5: Basic approach regarding BioConcolving [17]. The original signal is randomly splited into a predefined number of small parts. Those parts are then convolved between each ohter. Current figure was extracted from [17].

This parameter is then applied to the original signals, thus generating shifted versions of them. These approaches can be applied individually or simultaneously in order to increase the total number of possible convolutions. They also performed an invertibility analysis. Their premise was that an attacker had information about two different sequences generated from the same original signal. They concluded that to solve the deconvolution is as hard as random guessing. When the method was applied to on-line signature, they observed just a slight loss of performance regarding EER, for both approaches described above.

3.2.1 Applying Cancelable Biometrics to ECG

The application of cancelable biometrics to ECG is residual. One of the already proposed methods was developed by Dey et.al [6]. Their approach was based in biohashing and can be described in the following way (see figure 3.6):

1. A modified version of Pan-Tompkins algorithm was applied in order to detect P and T waves and

QRS complex.

2. The following features (time durations) were extracted: R-R, S-S, Q-Q, T-T, P-R, Q-T, Q-Tc and QRS complex
3. A tokenized number is randomly generated.
4. An inner product is performed between the matrix containing the extracted features and the tokenized number.
5. Finally, a biohash code is generated by comparing the inner product result with a predefined threshold.

Thus, the biohash code can be used to perform authentication of the subjects. According to them, this method is less susceptible to noise and more resistant to intra-class similarity, however the database used was small and it would be useful to apply this methodology to a larger set of records in order to prove its performance. Besides that, they do not present results regarding authentication performance.

The work of Dey et al. [6] will be used as a way to have a comparison regarding the method proposed by this work. Some changes on the methodology reported above were made, namely:

1. An algorithm provided by the group where the present work is included was applied to the heartbeats in order to detect P, R and P waves.
2. The features used to built the matrix were the following time intervals: P-P, R-R, T-T, P-R, P-T and R-T.
3. Authentication was performed by applying the methodology mentioned above to a larger database and the EER was computed.

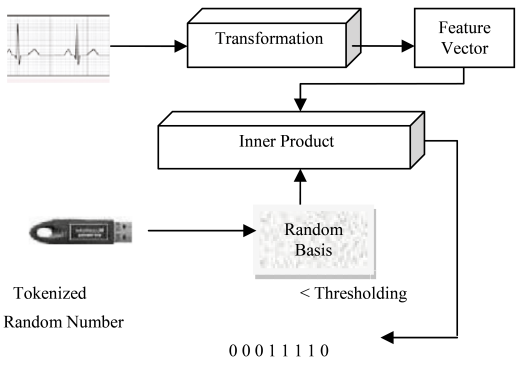

Figure 3.6: Proposed method by Dey et.al. [6]. Current figure was extracted from [6].

Chapter 4

Proposed Methodology

The main purpose of the present work is to develop a new cancelable biometric method by using ECG as biometric modality. To do that, two major study fields were explored in order to fulfil the cancelable templates criteria.

On the one hand, hermite polynomial approximation was applied to heartbeats, modelling the QRS complexes. This process ensures *irreversibility* and *unlikability* as explained below on the current chapter. On the other hand, a dissimilarity based representation was applied to the heartbeats' hermite approximations, which also guarantees *irreversibility* and *unlinkability*, as explained below too. Additionally, this representation is also applied to the original heartbeats, since, by itself, it leads to a cancelable biometric system.

The main system's blocks are depicted below (see figure 4.1):

1. The system's input is an ECG regarding lead I. The database used to test the developed system contains already filtered signals. Signals were filtered using a FIR (order 150) and with cut-off frequencies of 5Hz and 20Hz. More information regarding the database used can be found in section 5.1.

2. Filtered signals are then divided into heartbeats by a process of segmentation based on the work of Hamilton [11].

3. Hermite polynomial approximation is applied to model the QRS complex from each heartbeat.

4. The previous result is then used to feed a dissimilarity based classifier. The classifier will be tested in both situations: authentication and identification.

Note that the steps described above concern the case when the dissimilarity based classifier uses as input the hermite polynomial approximations. When the classifier uses as input the original heartbeats, step 3 is omitted. The remaining of the process is the same.

The results related to this approach will be compared with a traditional biometric system (section 3.1.1), and with a modified version of the cancelable ECG biometric system depicted in section 3.2.1.

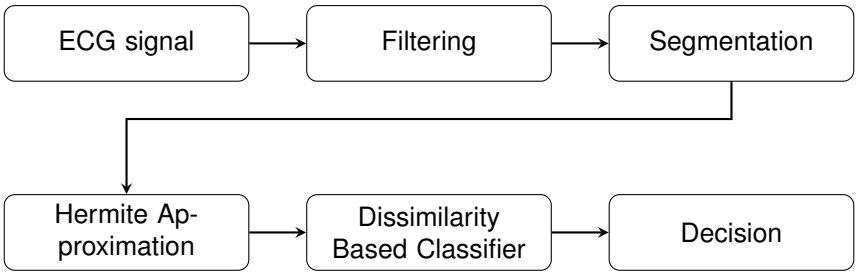

Figure 4.1: Block diagram regarding the proposed methodology, when the classifier uses as input the hermite polynomial approximations.

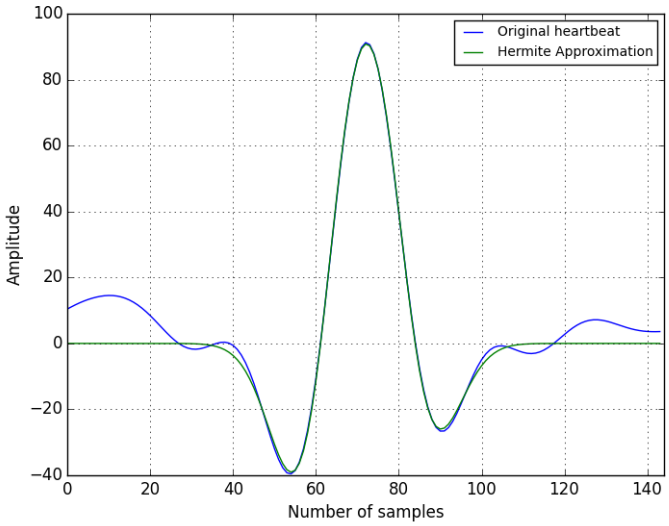

Figure 4.2: Original heartbeat (blue) and its hermite polynomial approximation (green), when 6 hermite functions are used to model the QRS complex.

4.1 Hermite Polynomial Approximation

As explained in section 2.6, hermite polynomial approximations can be used as a way to obtain a different representation of the QRS complex (see figure 4.2). This approximation is done by minimizing equation 2.12. Thus, it is not possible the recovery of the original QRS complex from its approximation, meaning that this process guarantees *irreversibility*. Additionally, since one can choose the number of polynomials to perform the approximation and the minimization can converge to a different minimum of eq 2.12, this process ensures *unlikability*. Thus, in the event of the system being compromised, it is always possible to generate different approximations.

In the present work, a 200 ms window was centred around each R peak. The window chosen is wide enough to cover all the QRS complex, but narrow enough to not include P and T waves [18]. Since all the hermite functions converges to zero in $\pm\infty$, it was added 100ms on each side of the QRS window. Thus, the total window's length is 400ms.

In order to observe the information loss in the hermite polynomial approximation, an error measure was introduced: the normalized root-mean square error (NRMSE). This error measure is defined as

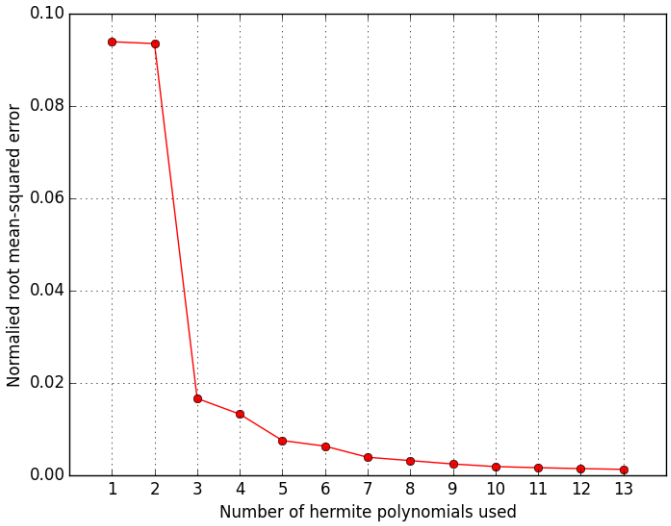

Figure 4.3: Normalized root mean-squared error

follows:

$$NRMSE = \frac{RMSE}{x_{max} - x_{min}} = \frac{\frac{\sqrt{\sum_t |e(t)|^2}}{N}}{x_{max} - x_{min}} \tag{4.1}$$

where N is the size of window in samples, the denominator represents the signals range of values and $|e(t)|$ is the difference between the original signal and its approximation. This quantity can be seen as the average error expressed as a percentage of the range of values in the signal fragment $(x_{max} - x_{min})$ [18].

As shown in Figure 4.3, NRMSE is higher if the approximation uses just one or two hermite polynomials and there is no significant improvement if the algorithm uses eight or more polynomials. Besides that, the computational time required to perform approximations grows proportionally with the number of polynomials used. Thus, in the present work, approximations from 4 to 7 polynomials are used across several experiments, described in the next chapter.

4.2 Dissimilarity Based Representation and Classification

As mentioned in section 2.7, a dissimilarity based representation computes the dissimilarity between an object and a set of prototypes, resulting in vector which is used by a classifier to perform a decision. In the present work, the objects which serve as input to the classifier are, on the one hand, the ECG heartbeats, and on the other hand, its approximations resulting of the hermite polynomial approximation process. As explained above, the approximation process covers the criteria used to design cancelable templates. However, when using as input to the classifier the original heartbeats it is also possible to obtain cancelable templates, as explained on section 4.3.1.

4.2.1 Notation

To simplify the understanding of the classifier, some notation that will be used in the rest of this work is presented:

- A population of S subjects;
- A testing population denoted by S_{tst};
- A training population denoted by S_{tr};
- A percentage per representing a fraction of the population which will be denoted as population S_{per}, randomly chosen from S_{tr};
- h_i representing a heartbeat associated to the subject i with $i = 1, ..., S$;
- H_i representing the total number of heartbeats associated to each subject i with $i = 1, ..., S$;
- a_i^p is a hermite polynomial approximation associated to the subject i, with $i = 1, ..., S$ and p represents the number of polynomials used with $p \in [3, 7]$;
- A_i^p is the total number of hermite polynomial approximations from a subject i, with $i = 1, ..., S$ and p represents the number of polynomials used with $p \in [3, 7]$;
- The generated dissimilarity space, D_S either using H_i or A_i^p.

4.3 Dissimilarity Representation

The approach followed to extract dissimilarities is described in section 4.3 and it is named **inter-subject approach**. In contrast to the work developed by [9], the system will be applied only to one lead.

The **inter-subject approach** can be explained as follows: the dissimilarity space is built based on a randomly chosen population, S_{per}. S_{per} is composed by a certain number of samples, representative of the population, called *prototypes*. For the tests described on chapter 5, $per \in [10, 90]\%$ in 10% intervals.

Two different dissimilarity representations are built regarding the proposed system: one at the time of the enrolment and another regarding the moment of classification.

4.3.1 Classifier Description

In order to better understand the proposed system, its modes are depicted next:

- **Enrolment** (see figure 4.5)
    1. Following a simple random sampling approach, a S_{tr} is built by choosing five samples regarding each subject, for all the subjects. Hereafter the samples belonging to S_{tr} will be called *templates*;

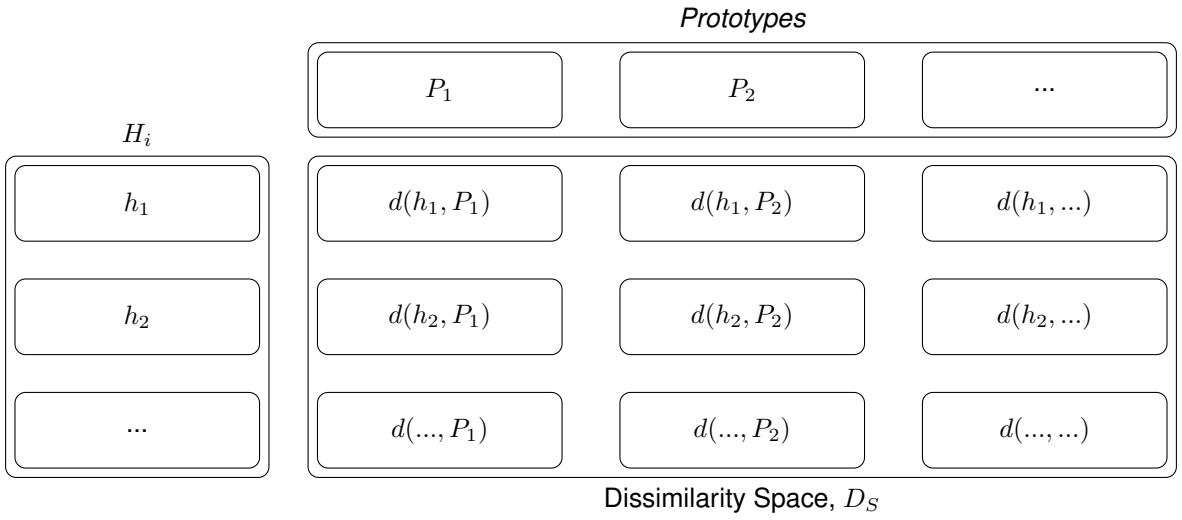

Figure 4.4: Dissimilarity space generation. In this example, the system uses as input the original records. If one considers H_i as being the templates regarding the i^{th} subject, thus D_S represents the dissimilarity space which is built at the enrolment phase.

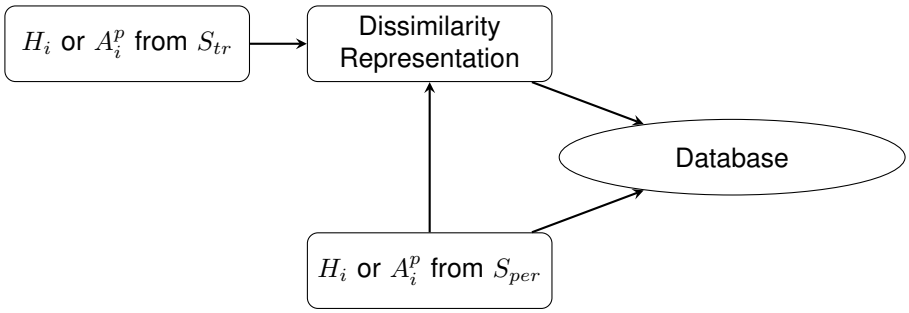

Figure 4.5: Enrolment regarding the dissimilarity based classifier.

2. Based on S_{tr}, a percentage per of subjects is randomly chosen leading to a sub population S_{per}. As referred above, the samples composing S_{per} are named *prototypes*;
3. For all H_i or A_i^p from S_{tr}, a dissimilarity representation is built. In other words, a dissimilarity representation of each template is represented by a dissimilarity vector of that template to each of the prototypes. Thus, a D_S is obtained (see figure 4.4).
4. D_S and H_i or A_i^p from S_p are stored.

- **Authentication/Identification** (see figure 4.6)

1. The samples not belonging to S_{tr} are used as test samples, meaning they define the S_{tst};
2. For all H_i or A_i from D_{tst}, a dissimilarity representation is built. This step is similar to step 2 of the enrolment phase. Another D_S is built.
3. A decision is made, taking into account the evaluation performed by a k-NN classifier by comparing the two generated dissimilarity representations from S_{tr} and S_{tst}.

Since it is possible to set per (note that this is a random process between the heartbeats or approximations of a given subject and dataset), in case of the system becoming compromised, it is possible

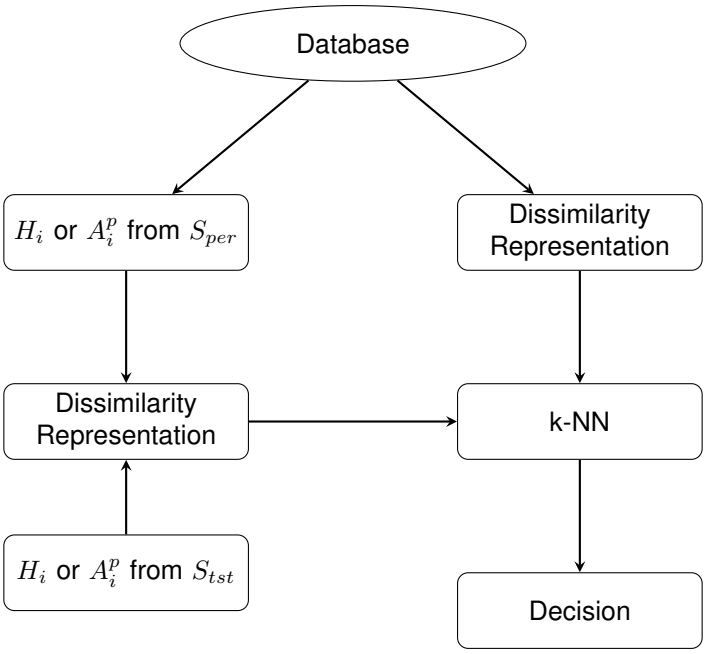

Figure 4.6: Recognition process regarding the dissimilarity based classifier.

to generate different templates, ensuring **renewability**. Besides that, the system is working on a dissimilarity space, meaning it is not possible to get back to the heartbeat/approximation space since the same dissimilarity representation can be obtained by several different pairs of objects, this process also guarantees *irreversibility*.

The proposed approach uses two measures: one serves as dissimilarity measure to compute the dissimilarity representation while the another serves as classification metric. Besides that, one can choose any measure since the properties mentioned in section 2.7.1 are met. As said on section 2.7.1, two different measures were considered: euclidean distance and cosine similarity. Those two measures allow the possibility of different performance tests:

- Cosine similarity used as both dissimilarity measure and classification metric, labelled as **CC**.
- Cosine similarity used as dissimilarity measure and euclidean distance used as classification metric, labelled as **CE**.
- Euclidean distance used as both dissimilarity measure and classification metric, labelled as **EE**.
- Euclidean distance used as dissimilarity measure and cosine similarity used as classification metric, labelled as **EC**.

In order to evaluate the system's performance, two different measures were calculated. On the one hand, regarding authentication, EER was computed taking into account both FAR and FRR regarding each test. One the another hand, PC_{ID} was computed based on Pe_{ID} and RR_{ID}, in order to evaluate the system's performance in the identification mode. Both authentication and identification measures were computed based on a predefined set of thresholds, and all the presented values in chapter 5 are with respect to the performance regarding all the subjects of the used database.

Chapter 5

Results

5.1 HSM Database

The HSM database was employed to apply the proposed methodology.

The records belonging to this database were collected at Santa Marta Hospital during normal hospital operation, meaning scheduled appointments, emergency cases and bedridden patients. The signals were acquired using Philips PageWriter Trim III devices, with a sampling rate of 500Hz and 16bit resolution.

Each record is composed by the records regarding the 12 leads and have a duration of about 10s. Since only the records belonging to healthy individuals (in respect to heart diseases) were considered, the following tests were applied to 832 records belonging to 612 subjects.

5.2 Genuine and Impostor Distributions

As said on section 2.7.1, the two dissimilarity measures that were considered were the euclidean distance and the cosine similarity. To make sure that the system would use the best combination possible, a series of tests were made. These tests aimed to obtain genuine and impostor distributions and to compute the superposition area between both distributions. Thus, the four combinations outlined in section 4.3.1 were tested.

These tests were ran by applying a simple random sampling approach. All the heartbeats were grouped according to the subject they belong to. Next, for each individual, five of them were randomly chosen to serve as templates to the classifier. The remaining beats were used individually to compute the distances in order to generate both distributions. As explained in section 4.3, dissimilarity based classifier performs its decision on the dissimilarity space, therefore, before the actual computation of distributions, a set of dissimilarity representations, D_S, were computed from the templates and the remaining beats, according to the process described in section 4.3.1. Finally, it was possible to generate the distributions, in the following way (see figure 5.1):

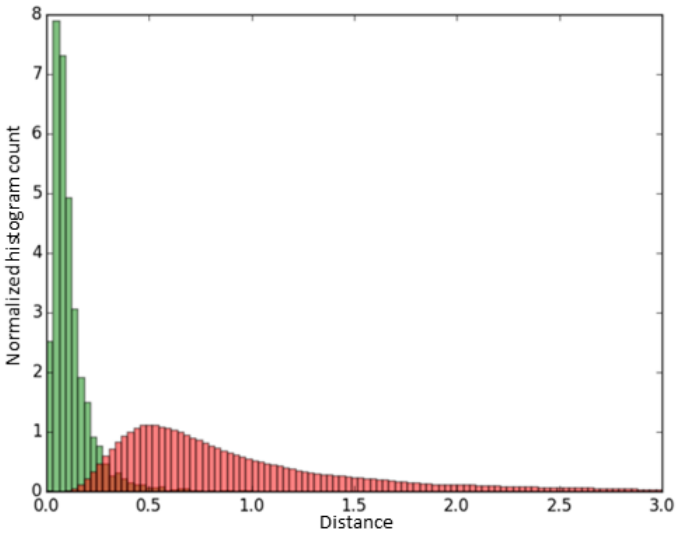

Figure 5.1: An example of the obtained genuine and impostor distributions regarding the original records, with $per = 10\%$. In this particular case, the distributions were obtained using as dissimilarity measure the cosine similarity and as classification metric the euclidean distance. The genuine and impostor distributions are represented respectively in green and red.

- By taking the distances between dissimilarity representations from a given subject and its templates, it was possible to generate the genuine distribution.
- The distances between the dissimilarity representation of a given subject and the templates regarding the other subjects were used to generate the impostor distribution.
- This approach was applied to all the subjects from the HSM database. Besides application to the original heartbeats, the hermite polynomial approximations also were considered.

The obtained histograms were normalized, meaning its integral area is 1. The percentage of population used, per, in each classifier varied from $[10, 90]\%$ in 10% intervals. This process was repeated ten times, in order to obtain the average superposition area as a function of the per used by the classifier. For the sake of simplicity, only the average superposition area curves of the several dissimilarity representations are plotted for the four combinations in figure 5.2. The standard deviation values regarding the distribution's superposition area are depicted in appendix A.

As seen in Figure 5.2, it is not clear what was the best combination, meaning the genuine and impostor distributions cannot be used as unique criteria to decide clearly what was the best pair. Thus, in order to clarify what pair was the best, the four combinations were tested by applying them to the recognition system presented in section 5.3.1. Two different performance parameters were computed: EER, related to the authentication process; and PC_{ID}, regarding identification. The approach followed was similar to the described above regarding the genuine/impostor distributions. The records were grouped according to the subject they belong to and 5 heartbeats were randomly chosen as templates, while the remaining ones were individually used to test the system. The number of neighbours, k, was set to 1 and the system was ran for the values of per on the range of $[10, 90]\%$ in 10% intervals. This process was repeated five times and it was applied to the original heartbeats as well as its approximations. To allow

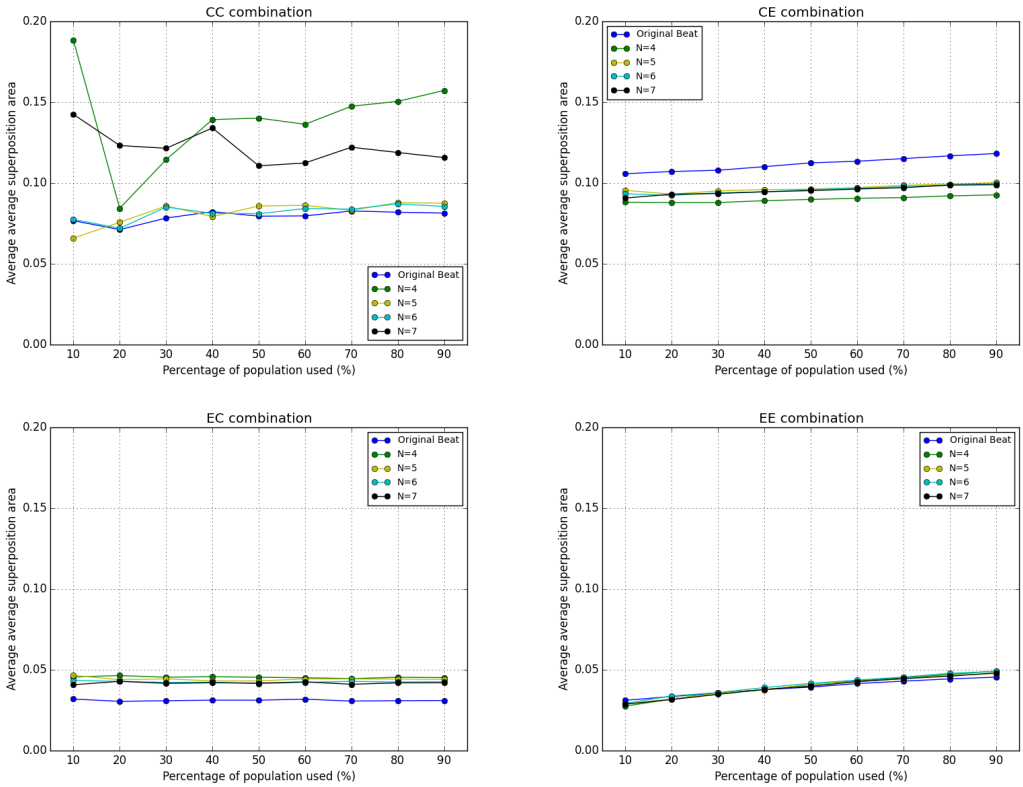

Figure 5.2: The influence of the percentage of population on the 4 combinations.

a better visualization of the results, only EER and PC_{ID} average values were plotted in figures 5.3, 5.4, 5.5 and 5.6. Standard deviation values can be observed from table 5.1 to table 5.8.

By comparing the previous results and figure 5.2 it was possible to conclude that the combination which was able to accomplish best results was EC combination. This combination was then used to perform some more tests on the system, which are depicted below on the next sections of this chapter.

5.3 Biometric Analysis

As said above, the EC combination was chosen to perform the remaining tests on the proposed method. Besides being different tests, all of them have some properties in common, namely, all of them

Table 5.1: Computed **EER** average and standard deviation values, regarding **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	4.61 $\pm$ 0.37	6.75 $\pm$ 0.10	6.47 $\pm$ 0.29	6.25 $\pm$ 0.39	6.16 $\pm$ 0.26
20	4.63 $\pm$ 0.44	6.81 $\pm$ 0.31	6.50 $\pm$ 0.22	6.12 $\pm$ 0.23	6.33 $\pm$ 0.15
30	4.59 $\pm$ 0.36	6.70 $\pm$ 0.29	6.46 $\pm$ 0.22	6.17 $\pm$ 0.27	6.23 $\pm$ 0.16
40	4.64 $\pm$ 0.37	6.70 $\pm$ 0.25	6.33 $\pm$ 0.17	6.14 $\pm$ 0.22	6.21 $\pm$ 0.20
50	4.53 $\pm$ 0.32	6.65 $\pm$ 0.29	6.35 $\pm$ 0.15	6.16 $\pm$ 0.24	6.21 $\pm$ 0.14
60	4.54 $\pm$ 0.35	6.65 $\pm$ 0.30	6.37 $\pm$ 0.19	6.07 $\pm$ 0.21	6.16 $\pm$ 0.14
70	4.60 $\pm$ 0.36	6.67 $\pm$ 0.30	6.33 $\pm$ 0.22	6.13 $\pm$ 0.23	6.17 $\pm$ 0.20
80	4.56 $\pm$ 0.34	6.64 $\pm$ 0.29	6.33 $\pm$ 0.17	6.14 $\pm$ 0.22	6.20 $\pm$ 0.16
90	4.58 $\pm$ 0.34	6.65 $\pm$ 0.28	6.36 $\pm$ 0.16	6.13 $\pm$ 0.22	6.20 $\pm$ 0.16

Table 5.2: Computed PC_{ID} average and standard deviation values regarding **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	12.27 ± 0.65	38.18 ± 1.80	35.18 ± 2.45	33.85 ± 1.61	33.82 ± 1.64
20	11.93 ± 0.91	35.00 ± 1.53	34.10 ± 1.66	33.99 ± 1.55	33.91 ± 1.46
30	11.96 ± 0.39	35.09 ± 1.43	34.05 ± 1.49	33.81 ± 1.53	33.75 ± 1.53
40	11.93 ± 0.72	35.99 ± 2.49	33.96 ± 1.50	33.69 ± 1.54	33.60 ± 1.61
50	11.80 ± 0.42	34.92 ± 1.56	33.86 ± 1.63	33.71 ± 1.52	33.63 ± 1.46
60	11.72 ± 0.54	34.94 ± 1.53	35.90 ± 1.65	33.79 ± 1.54	33.54 ± 1.61
70	11.71 ± 0.40	34.90 ± 1.52	33.89 ± 1.66	33.70 ± 1.54	33.53 ± 1.62
80	11.85 ± 0.28	34.97 ± 1.52	33.86 ± 1.68	33.67 ± 1.53	33.64 ± 1.53
90	11.85 ± 0.26	34.95 ± 1.50	33.89 ± 1.65	33.64 ± 1.53	33.60 ± 1.58

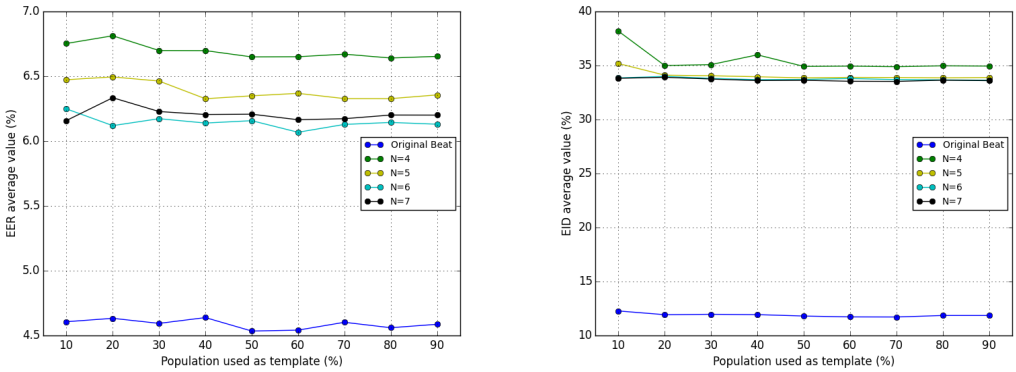

Figure 5.3: Computed **EER** and PC_{ID} average values regarding **EC** combination.

Table 5.3: Computed **EER** average and standard deviation values regarding the **EE** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	5.58 ± 0.19	7.14 ± 0.10	6.99 ± 0.27	6.82 ± 0.30	7.01 ± 0.41
20	5.51 ± 0.25	7.33 ± 0.14	6.94 ± 0.11	6.73 ± 0.21	6.86 ± 0.19
30	5.41 ± 0.23	7.25 ± 0.11	6.97 ± 0.20	6.71 ± 0.24	6.85 ± 0.29
40	5.39 ± 0.27	7.27 ± 0.11	6.99 ± 0.19	6.72 ± 0.20	6.80 ± 0.22
50	5.42 ± 0.30	7.26 ± 0.05	6.96 ± 0.19	6.73 ± 0.21	6.80 ± 0.21
60	5.51 ± 0.29	7.26 ± 0.08	6.90 ± 0.25	6.70 ± 0.21	6.74 ± 0.23
70	5.36 ± 0.28	7.19 ± 0.06	6.91 ± 0.23	6.70 ± 0.24	6.75 ± 0.25
80	5.31 ± 0.28	7.16 ± 0.07	6.88 ± 0.22	6.72 ± 0.24	6.76 ± 0.27
90	5.36 ± 0.24	7.15 ± 0.03	6.85 ± 0.22	6.71 ± 0.22	6.76 ± 0.26

Table 5.4: Computed PC_{ID} average and standard deviation values regarding **EE** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	15.03 ± 1.00	39.65 ± 1.24	38.59 ± 1.18	38.24 ± 1.23	38.47 ± 1.04
20	14.94 ± 0.65	38.46 ± 1.75	36.72 ± 0.37	36.16 ± 0.57	35.90 ± 0.74
30	16.25 ± 0.71	39.21 ± 1.18	38.12 ± 0.55	37.45 ± 0.68	37.33 ± 0.60
40	15.23 ± 0.30	39.77 ± 1.17	37.89 ± 1.80	36.48 ± 1.84	35.27 ± 1.32
50	16.03 ± 0.54	38.14 ± 20.3	36.49 ± 0.61	35.98 ± 0.66	35.88 ± 0.88
60	15.24 ± 0.27	38.56 ± 1.65	36.97 ± 0.27	36.60 ± 0.37	36.42 ± 0.33
70	15.29 ± 1.18	36.84 ± 2.45	37.56 ± 0.34	37.03 ± 0.18	36.81 ± 0.40
80	14.65 ± 0.90	37.25 ± 2.57	36.87 ± 2.73	36.33 ± 2.67	37.31 ± 0.70
90	15.14 ± 1.03	37.51 ± 2.84	35.53 ± 2.13	35.05 ± 1.98	34.82 ± 1.76

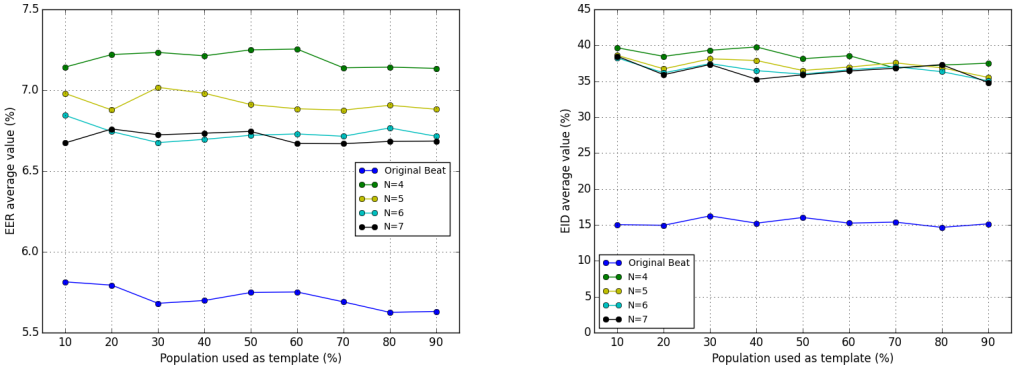

Figure 5.4: Computed **EER** and PC_{ID} average values regarding **EE** combination.

Table 5.5: Computed **EER** average and standard deviation values regarding **CC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	4.82 ± 1.95	11.54 ± 1.37	14.10 ± 2.47	11.23 ± 1.64	10.65 ± 1.59
20	4.91 ± 1.54	13.06 ± 1.11	14.17 ± 0.54	12.18 ± 1.42	12.67 ± 1.82
30	5.86 ± 0.88	13.22 ± 0.43	12.63 ± 1.22	11.34 ± 1.32	10.98 ± 1.72
40	6.34 ± 0.31	12.65 ± 1.16	13.35 ± 1.07	12.00 ± 0.87	12.55 ± 0.69
50	5.63 ± 0.50	13.28 ± 0.43	13.33 ± 0.94	12.27 ± 0.39	12.60 ± 0.47
60	5.90 ± 0.32	13.38 ± 0.28	13.35 ± 0.70	12.35 ± 0.41	12.61 ± 0.30
70	5.89 ± 0.51	13.49 ± 0.35	13.09 ± 0.58	12.22 ± 0.29	12.66 ± 0.43
80	5.90 ± 0.46	13.35 ± 0.30	13.42 ± 0.35	12.50 ± 0.40	12.72 ± 0.48
90	5.86 ± 0.23	13.56 ± 0.16	13.40 ± 0.20	13.33 ± 0.60	12.71 ± 0.46

Table 5.6: Computed PC_{ID} average and standard deviation values regarding **CC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	7.97 ± 0.99	42.64 ± 0.56	41.79 ± 0.42	41.53 ± 0.70	40.33 ± 1.41
20	8.39 ± 0.99	43.42 ± 0.77	41.88 ± 0.53	41.45 ± 0.72	41.19 ± 0.95
30	8.59 ± 0.63	42.92 ± 0.51	42.03 ± 0.63	41.56 ± 1.03	41.40 ± 0.74
40	8.81 ± 0.46	43.06 ± 0.42	42.22 ± 0.53	41.92 ± 0.53	41.31 ± 0.43
50	8.57 ± 0.24	42.85 ± 0.38	41.94 ± 0.23	41.83 ± 0.60	41.80 ± 0.69
60	8.64 ± 0.18	43.05 ± 0.62	42.01 ± 0.41	41.93 ± 0.62	41.82 ± 0.61
70	8.76 ± 0.67	43.02 ± 0.55	41.83 ± 0.49	41.93 ± 0.78	41.92 ± 0.68
80	8.76 ± 0.48	43.01 ± 0.67	42.16 ± 0.68	41.78 ± 0.66	41.77 ± 0.64
90	8.63 ± 0.46	42.88 ± 0.50	42.09 ± 0.46	41.44 ± 0.51	41.80 ± 0.64

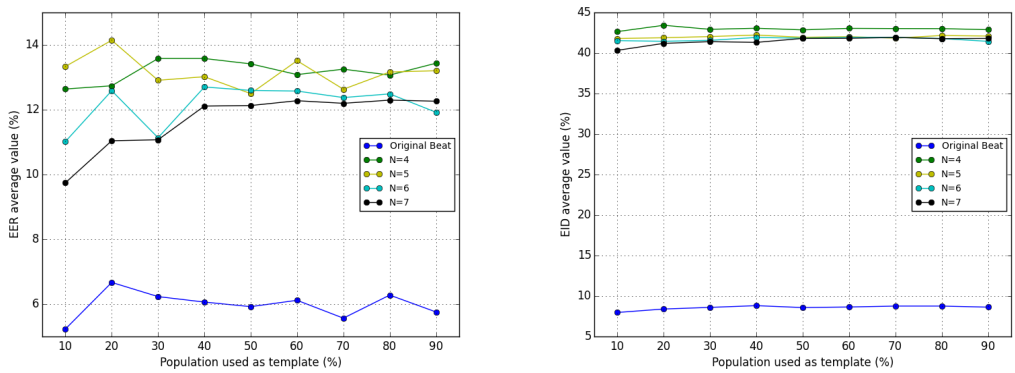

Figure 5.5: Computed **EER** and PC_{ID} average values regarding **CC** combination.

Table 5.7: Computed **EER** average and standard deviation values regarding **CE** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	9.24 ± 0.32	14.76 ± 0.86	14.63 ± 0.98	13.93 ± 0.85	13.35 ± 0.39
20	8.63 ± 0.39	14.26 ± 0.50	13.92 ± 0.29	13.59 ± 0.26	13.13 ± 0.56
30	8.81 ± 0.42	14.26 ± 0.42	14.15 ± 0.47	13.62 ± 0.58	13.20 ± 0.63
40	8.75 ± 0.14	14.03 ± 0.44	14.15 ± 0.52	13.48 ± 0.52	12.99 ± 0.62
50	8.58 ± 0.29	14.25 ± 0.45	13.92 ± 0.35	13.27 ± 0.26	13.07 ± 0.61
60	8.39 ± 0.27	13.98 ± 0.37	13.92 ± 0.35	13.29 ± 0.20	13.03 ± 0.74
70	8.73 ± 0.04	13.96 ± 0.36	13.97 ± 0.35	13.32 ± 0.35	12.94 ± 0.74
80	8.42 ± 0.40	13.99 ± 0.35	13.92 ± 0.32	13.31 ± 0.27	13.00 ± 0.69
90	8.39 ± 0.24	13.87 ± 0.31	13.90 ± 0.29	13.33 ± 0.26	13.04 ± 0.74

Table 5.8: Computed PC_{ID} average and standard deviation values regarding **CE** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	11.40 ± 0.76	44.14 ± 0.76	42.40 ± 0.23	41.70 ± 0.32	39.94 ± 0.39
20	13.09 ± 0.42	44.02 ± 0.27	42.66 ± 0.24	42.62 ± 0.63	40.49 ± 0.33
30	10.34 ± 0.26	43.91 ± 0.10	43.08 ± 0.58	42.53 ± 0.43	40.56 ± 0.37
40	9.93 ± 0.15	44.48 ± 0.15	43.54 ± 0.28	43.25 ± 0.30	40.84 ± 0.38
50	10.10 ± 0.64	44.19 ± 0.36	43.34 ± 0.66	43.01 ± 0.62	40.99 ± 0.42
60	10.40 ± 0.35	44.42 ± 0.14	43.69 ± 0.49	43.36 ± 0.41	40.68 ± 0.59
70	10.00 ± 0.31	44.25 ± 0.45	43.51 ± 0.81	43.32 ± 0.70	40.09 ± 0.43
80	10.03 ± 0.47	44.52 ± 0.39	43.84 ± 0.74	42.05 ± 0.97	39.16 ± 0.54
90	10.17 ± 0.36	44.76 ± 0.40	42.30 ± 0.56	42.10 ± 0.73	39.50 ± 0.55

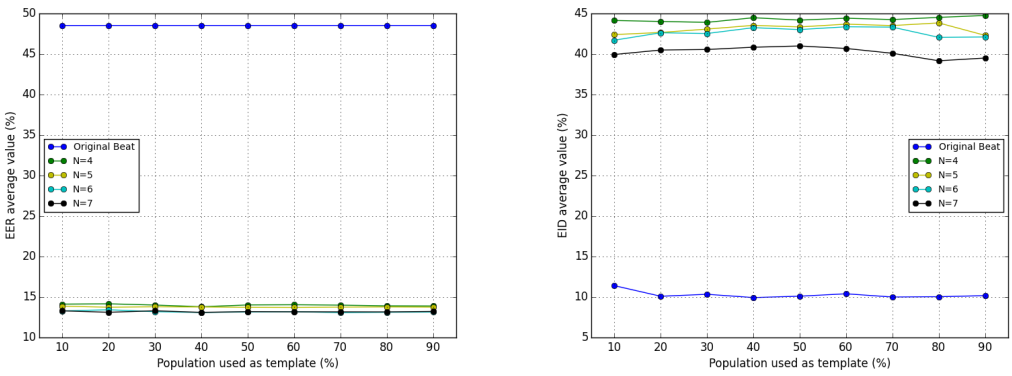

Figure 5.6: Computed **EER** and PC_{ID} average values regarding **CE** combination.

used 5 randomly chosen heartbeats or hermite polynomial approximations regarding each subject in order to produce the training set, S_{tr}. The remaining data related with a given individual was used to built the test set, S_{tst}. At the recognition phase, each heartbeat or approximation was tested individually. All of the tests presented below were ran 5 times each. EC metric combination was used in all of the tests presented on the remaining of this section.

Besides that, a series of tests were made regarding the biometric system described on sections 3.1.1 and 3.2.1 in order to turn possible a comparison between the proposed method's performance and the performance obtained by those recognition systems.

Table 5.9: Computed **EER** average and standard deviation values regarding the proposed method (**approach 2**, k=1), using **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	2.55 $\pm$ 0.05	5.13 $\pm$ 0.05	4.95 $\pm$ 0.07	4.80 $\pm$ 0.07	4.78 $\pm$ 0.21
20	2.60 $\pm$ 0.10	4.85 $\pm$ 0.17	4.84 $\pm$ 0.04	4.84 $\pm$ 0.19	4.79 $\pm$ 0.16
30	2.63 $\pm$ 0.05	4.99 $\pm$ 0.03	4.85 $\pm$ 0.04	4.87 $\pm$ 0.15	4.75 $\pm$ 0.18
40	2.64 $\pm$ 0.10	4.96 $\pm$ 0.02	4.90 $\pm$ 0.13	4.81 $\pm$ 0.16	4.75 $\pm$ 0.15
50	2.67 $\pm$ 0.09	4.91 $\pm$ 0.08	4.88 $\pm$ 0.04	4.84 $\pm$ 0.16	4.71 $\pm$ 0.16
60	2.65 $\pm$ 0.07	4.89 $\pm$ 0.08	4.85 $\pm$ 0.09	4.78 $\pm$ 0.15	4.75 $\pm$ 0.14
70	2.66 $\pm$ 0.03	4.96 $\pm$ 0.06	4.90 $\pm$ 0.01	4.81 $\pm$ 0.10	4.76 $\pm$ 0.13
80	2.62 $\pm$ 0.04	4.91 $\pm$ 0.09	4.88 $\pm$ 0.10	4.84 $\pm$ 0.12	4.77 $\pm$ 0.14
90	2.61 $\pm$ 0.09	4.96 $\pm$ 0.10	4.90 $\pm$ 0.07	4.85 $\pm$ 0.11	4.75 $\pm$ 0.20

5.3.1 Proposed System

Since the database used has for several subjects different record sessions, it was considered to be interesting to test the proposed methodology with four different approaches in order to observe the intra-variability degree between different sessions belonging to the same subject, thus evaluating system's performance under different conditions. The four used approaches were:

- **Approach 1** - Records were grouped by the subjects they belong to. Thus, the number of tested subjects was 612. No distinction between sessions was taken into account at the moments of training or testing. This approach was the one used to set the best metric combination metric described on section 5.2.
- **Approach 2** - Each of the records were considered to be a different subject, meaning that the total number of tested "subjects" was 832.
- **Approach 3** - Records were grouped by the subjects they belong to but only one record session was used at the training phase, while the remaining were used at the test phase.
- **Approach 4** - Records were grouped by the subjects they belong to but in contrast to approach 3, training was performed with several sessions and only one session was used to evaluate the system's performance.

To evaluate the system's performance, two different measures were computed: on one hand, EER was computed regarding the authentication process and one the other hand, PC_{ID} was computed relative with the identification. Both of the measures were computed by evaluating the system's performance while working under a threshold set.

Since the results related with **approach 1** are depicted in tables 5.1 and 5.2 and in figure 5.3, only the results regarding **approach 2**, **approach 3** and **approach 4** are depicted below. Those tests were ran using k set to 1.

In order to observe also the influence of classifier's neighbours, k, on the perform of the proposed system, this parameter was set to 3 and the system was tested using **approach 1**.

It was not completely clear that using a higher number of polynomials conducted to a better system's performance. Thus, since one must have present the computational time needed to perform the approx-

Table 5.10: Computed PC_{ID} average and standard deviation values regarding the proposed method (**approach 2**, k=1), using **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	13.38 ± 0.26	37.21 ± 2.90	39.97 ± 0.23	37.17 ± 2.17	36.84 ± 2.35
20	13.45 ± 0.12	39.82 ± 0.22	39.79 ± 0.28	39.00 ± 0.81	36.79 ± 2.28
30	13.65 ± 0.10	40.12 ± 0.16	37.07 ± 2.89	36.78 ± 2.66	36.74 ± 2.55
40	13.51 ± 0.01	40.08 ± 0.05	39.90 ± 0.05	38.91 ± 0.93	38.68 ± 0.98
50	13.53 ± 0.16	40.03 ± 0.04	36.93 ± 3.02	38.83 ± 1.03	38.78 ± 1.01
60	13.53 ± 0.06	40.14 ± 0.03	39.82 ± 0.01	38.93 ± 0.94	36.70 ± 2.70
70	13.51 ± 0.10	40.17 ± 0.01	36.89 ± 3.15	38.89 ± 0.95	36.73 ± 2.47
80	13.46 ± 0.04	40.13 ± 0.02	39.96 ± 0.13	38.89 ± 0.90	36.72 ± 2.72
90	13.59 ± 0.16	40.05 ± 0.01	39.91 ± 0.10	39.92 ± 0.93	38.77 ± 1.08

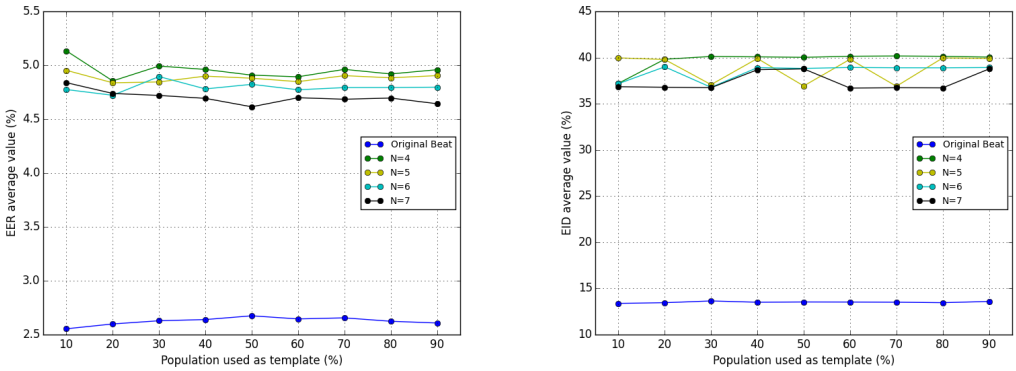

Figure 5.7: Computed **EER** and PC_{ID} average values regarding **approach 2** and k=1, using **EC** combination.

Table 5.11: Computed **EER** average and standard deviation values regarding the proposed method (**approach 3**, k=1), using **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	29.11 ± 0.44	29.23 ± 0.18	28.80 ± 0.55	28.11 ± 0.64	28.97 ± 0.34
20	28.61 ± 0.75	29.19 ± 0.72	29.08 ± 0.52	28.64 ± 0.54	28.63 ± 0.60
30	28.30 ± 0.99	29.33 ± 0.43	29.33 ± 0.43	28.64 ± 0.59	28.78 ± 0.78
40	28.60 ± 0.92	29.29 ± 0.32	29.55 ± 0.16	28.61 ± 0.57	28.92 ± 0.54
50	28.50 ± 0.93	29.20 ± 0.73	29.10 ± 0.52	28.56 ± 0.57	29.02 ± 0.75
60	28.56 ± 0.76	29.36 ± 0.32	29.05 ± 0.59	29.80 ± 0.36	28.82 ± 0.89
70	28.55 ± 0.92	29.39 ± 0.37	29.00 ± 0.54	28.80 ± 0.31	29.13 ± 0.61
80	28.47 ± 0.92	29.20 ± 0.66	29.28 ± 0.49	28.37 ± 0.71	29.09 ± 0.62
90	28.48 ± 0.91	29.35 ± 0.33	29.08 ± 0.53	28.45 ± 0.74	29.08 ± 0.59

Table 5.12: Computed PC_{ID} average and standard deviation values regarding the proposed method (**approach 3**, k=1), using **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	41.26 ± 0.19	43.93 ± 0.99	43.76 ± 1.17	43.25 ± 1.11	42.83 ± 0.90
20	41.15 ± 0.23	43.85 ± 0.94	43.25 ± 1.14	42.97 ± 1.27	43.30 ± 1.05
30	40.86 ± 0.22	43.42 ± 1.03	44.64 ± 0.11	43.21 ± 1.02	43.03 ± 1.14
40	40.87 ± 0.09	43.80 ± 1.10	43.68 ± 1.20	44.40 ± 0.12	43.54 ± 1.12
50	40.99 ± 0.20	43.81 ± 1.08	43.76 ± 0.93	42.64 ± 0.95	43.98 ± 0.92
60	40.77 ± 0.19	43.82 ± 1.11	43.73 ± 1.17	43.99 ± 0.83	43.98 ± 0.90
70	40.86 ± 0.20	43.81 ± 0.93	43.30 ± 1.15	43.95 ± 0.87	44.00 ± 0.90
80	40.78 ± 0.28	43.79 ± 1.03	43.73 ± 1.19	44.38 ± 0.08	43.95 ± 0.94
90	40.83 ± 0.27	43.33 ± 1.05	43.74 ± 1.14	43.97 ± 0.88	44.35 ± 0.07

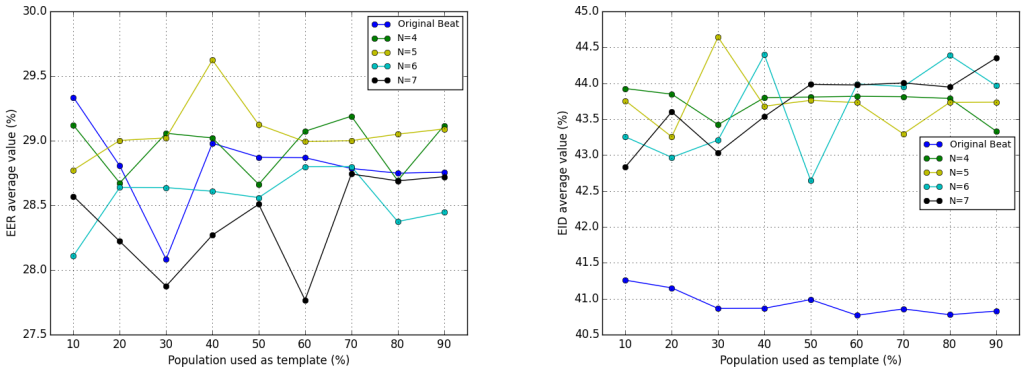

Figure 5.8: Computed **EER** and PC_{ID} average values regarding **approach 3** and k=1, using **EC** combination.

Table 5.13: Computed **EER** average and standard deviation values regarding the proposed method (**approach 4**, k=1), using **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	28.52 ± 1.26	28.88 ± 0.76	28.96 ± 0.51	28.48 ± 0.58	28.55 ± 0.73
20	28.63 ± 0.77	29.31 ± 0.61	28.99 ± 0.82	28.56 ± 0.88	28.24 ± 0.97
30	28.95 ± 1.19	29.09 ± 0.62	29.29 ± 0.53	28.67 ± 0.82	28.79 ± 0.93
40	28.59 ± 0.87	28.99 ± 0.85	29.11 ± 0.71	28.78 ± 0.61	28.61 ± 1.12
50	28.44 ± 0.89	28.95 ± 0.84	29.24 ± 0.47	28.47 ± 0.84	28.64 ± 0.95
60	28.47 ± 1.01	28.97 ± 0.85	29.24 ± 0.59	29.13 ± 0.43	28.94 ± 0.65
70	28.39 ± 1.01	29.14 ± 0.65	29.31 ± 0.35	29.05 ± 0.48	28.86 ± 0.82
80	28.48 ± 0.97	28.74 ± 0.83	29.07 ± 0.58	28.71 ± 0.45	28.82 ± 0.86
90	28.45 ± 0.97	29.00 ± 0.88	29.34 ± 0.37	28.95 ± 0.38	28.88 ± 0.81

Table 5.14: Computed PC_{ID} average and standard deviation values regarding the proposed method (**approach 4**, k=1), using **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	41.29 ± 1.24	42.70 ± 1.02	42.66 ± 1.11	43.19 ± 0.96	43.27 ± 0.90
20	41.62 ± 0.95	43.68 ± 1.12	43.02 ± 1.15	42.97 ± 1.02	42.89 ± 0.96
30	41.82 ± 1.45	43.13 ± 1.07	43.52 ± 1.11	42.94 ± 1.11	43.49 ± 0.97
40	40.53 ± 0.31	43.23 ± 0.98	43.58 ± 1.07	43.02 ± 0.99	42.96 ± 0.98
50	40.59 ± 0.20	43.64 ± 1.06	43.53 ± 1.16	42.91 ± 1.00	42.92 ± 1.02
60	41.17 ± 1.07	43.61 ± 1.07	43.50 ± 1.11	43.38 ± 1.07	43.36 ± 1.07
70	40.55 ± 0.20	43.25 ± 0.99	43.57 ± 1.11	43.39 ± 1.01	43.43 ± 0.99
80	40.60 ± 0.19	43.64 ± 1.01	43.08 ± 1.09	42.49 ± 0.83	43.45 ± 0.98
90	40.58 ± 0.15	43.63 ± 1.02	43.54 ± 1.13	42.93 ± 1.01	43.52 ± 0.98

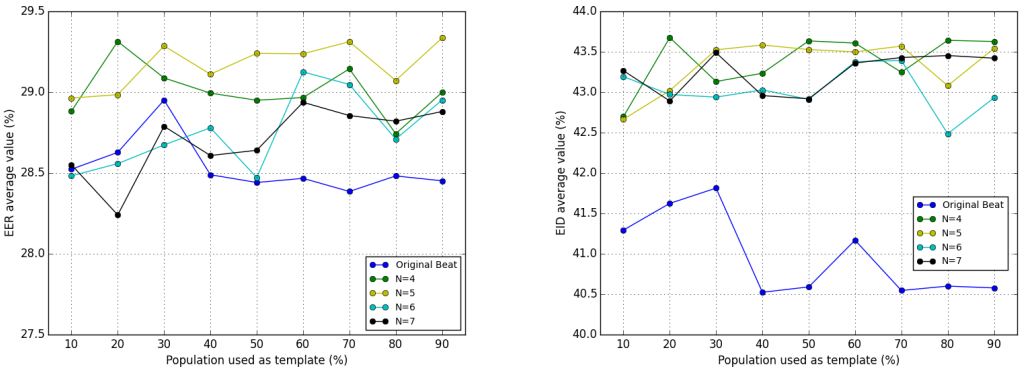

Figure 5.9: Computed **EER** and PC_{ID} average values regarding **approach 4** and k=1, using **EC** combination.

Table 5.15: Computed EER average and standard deviation values regarding the proposed method (**approach 1**, k=3), using **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	7.76 ± 0.24	9.38 ± 0.37	8.87 ± 0.18	8.55 ± 0.20	8.69 ± 0.15
20	7.73 ± 0.32	9.30 ± 0.37	7.77 ± 0.35	8.72 ± 0.33	8.72 ± 0.22
30	7.68 ± 0.16	9.27 ± 0.21	8.71 ± 0.22	8.70 ± 0.09	8.68 ± 0.19
40	7.71 ± 0.21	9.24 ± 0.25	8.73 ± 0.29	8.61 ± 0.18	8.65 ± 0.20
50	7.58 ± 0.19	9.26 ± 0.30	8.72 ± 0.14	8.68 ± 0.10	8.59 ± 0.12
60	7.63 ± 0.25	9.25 ± 0.23	8.71 ± 0.17	8.68 ± 0.21	8.55 ± 0.10
70	7.56 ± 0.17	9.25 ± 0.29	8.69 ± 0.20	8.66 ± 0.19	8.50 ± 0.15
80	7.55 ± 0.23	9.26 ± 0.24	8.71 ± 0.18	8.67 ± 0.11	8.48 ± 0.12
90	7.59 ± 0.22	9.31 ± 0.29	8.70 ± 0.19	8.66 ± 0.18	8.51 ± 0.12

Table 5.16: Computed PC_{ID} average and standard deviation values regarding the proposed method (**approach 1**, k=3), using **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	19.41 $\pm$ 1.11	39.96 $\pm$ 0.30	38.61 $\pm$ 1.06	38.93 $\pm$ 0.30	38.68 $\pm$ 0.11
20	18.73 $\pm$ 0.80	39.79 $\pm$ 0.48	38.58 $\pm$ 1.04	38.40 $\pm$ 0.97	38.84 $\pm$ 0.30
30	18.84 $\pm$ 0.93	39.83 $\pm$ 0.30	38.99 $\pm$ 0.42	38.25 $\pm$ 1.14	38.33 $\pm$ 1.02
40	18.62 $\pm$ 0.76	39.72 $\pm$ 0.28	38.99 $\pm$ 0.30	38.81 $\pm$ 0.34	38.76 $\pm$ 0.28
50	18.80 $\pm$ 1.06	39.78 $\pm$ 0.27	38.61 $\pm$ 1.04	38.80 $\pm$ 0.36	38.83 $\pm$ 0.23
60	18.79 $\pm$ 0.98	39.70 $\pm$ 0.29	38.96 $\pm$ 0.35	38.83 $\pm$ 0.35	38.74 $\pm$ 0.26
70	18.51 $\pm$ 0.90	39.69 $\pm$ 0.30	38.99 $\pm$ 0.36	38.76 $\pm$ 0.35	38.78 $\pm$ 0.17
80	18.52 $\pm$ 0.84	39.73 $\pm$ 0.21	39.02 $\pm$ 0.35	37.97 $\pm$ 1.15	38.81 $\pm$ 0.25
90	18.71 $\pm$ 0.95	39.77 $\pm$ 0.27	38.97 $\pm$ 0.36	38.79 $\pm$ 0.33	38.72 $\pm$ 0.19

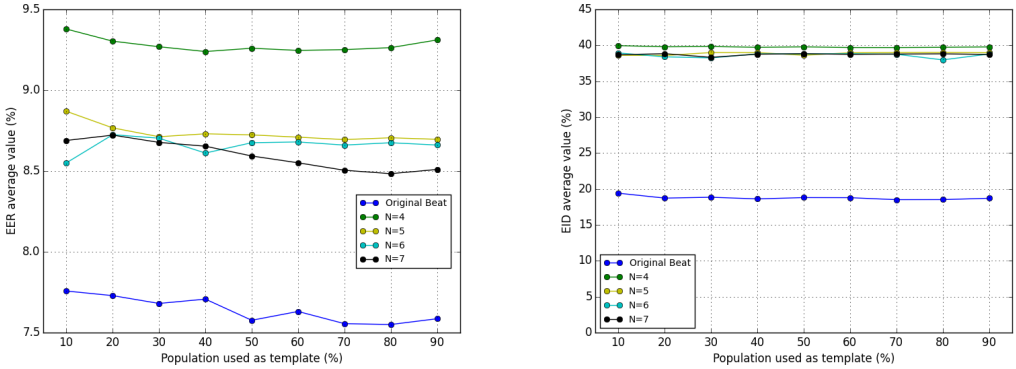

Figure 5.10: Computed **EER** and PC_{ID} average values regarding **approach 1** and k=3, using **EC** combination.

imation, in a real life situation, a small number of polynomials can be used, as long as the approximation error being reduced, thus saving time at enrolment and recognition phases. When comparing the approximations with the original records, it can be seen a decrease on system's performance, which was somewhat expected since heartbeats have more information available to be used in order to distinguish between the several subjects on a database.

Besides studying the influence of the parameters referred above, also the way of how database records were used was studied, as explained above. As expected, using **approach 2** produced better results, but one must take into account that this approach is just a fictional situation. In more real life situations, system presented a decrease in its performance. When using **approaches 3 and 4**, closer to a real life application, the system performance was not so good. However those results can be justified by the following facts: not all subjects had different record sessions and many of them had just two sessions recorded. This is an important point, since to produce a good training set it is useful to have as many records from the same subject as possible in order to cover its intra-variability.

In relation to k parameter, it was possible to conclude that system performed better with k set to one than with $k = 3$ in authentication mode as well as in identification mode. In relation to per, it was not completely clear which was the best per to use. In some tests, it is noted a slightly improvement associated to higher parameter values but there is not an established tendency. This analysis is true for both original records as well as its approximations.

Table 5.17: Computed **EER** and PC_{ID} values regarding the system proposed by Carreiras et al. [3].

	EER (%)	PC_{ID}**(%)**
Original beat	9.83 ± 0.30	10.90 ± 1.47
QRS complex	9.73 ± 0.36	33.25 ± 0.34
N = 4	13.14 ± 0.32	42.39 ± 0.40
N = 5	13.43 ± 0.51	41.19 ± 0.48
N = 6	12.16 ± 0.38	40.05 ± 0.28
N = 7	12.53 ± 0.38	39.78 ± 0.34

5.3.2 Biometric System Proposed by [3]

As referred above, the work of Carreiras et.al [3], described on section 3.1.1, was used in order to obtain a comparison regarding the proposed methodology.

In the scope of the present work, not only the heartbeats were used to train the classifier, but also its hermite polynomial approximations. The system was also applied to the QRS complex, therefore obtaining a better comparison term with the hermite polynomial approximations. To obtain the QRS complex, a 200ms window was centred around R peak, which is wide enough to cover all the QRS complex but narrow enough to not include P and T waves [18]. Thus, six different types of tests were done. For each one of them, a simple approach was done. The records were grouped by subject, similar to **approach 1** referred on section 5.3.1. For each subject, four heartbeats were randomly chosen, three of them were used as templates and the remaining one was used to test the classifier, according to [3]. This process was repeated 10 times for each classifier.

In order to evaluate the system's performance, EER and PC_{ID} were obtained. Table 5.17 presents the average and standard deviation values related with this system. Besides that, figure 5.11 presents ROC curve regarding each one of the made tests.

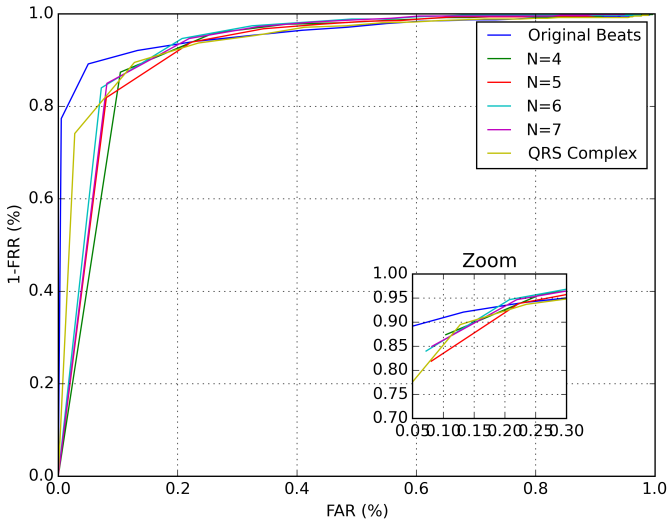

Figure 5.11: **ROC curve** for the hermite polynomial approximations and the original beats, regarding the system proposed by Carreiras et al. [3].

Since the followed approach is similar to **approach 1**, the obtained results can be compared to the

Table 5.18: Computed **EER** average and standard deviation values regarding the method proposed by Dey et al. [6], when **euclidean distance** is used as decision criterion

Discretization Threshold	EER (%)
0.16	27.00 ± 1.42
0.47	12.16 ± 0.95
0.79	10.86 ± 0.97
1.11	9.55 ± 0.47
1.42	10.07 ± 0.75
1.74	14.73 ± 5.48
2.05	29.73 ± 1.14
2.37	39.10 ± 1.09
2.68	44.07 ± 0.37
3	46.95 ± 0.64

ones depicted on tables 5.1 and 5.2 and figure 5.3. Those tests used the best combination (EC combination) by following **approach 1**. Thus, it is possible to conclude that the method presented in this work was able to achieve best results when it used the original records as well as when it used the hermite polynomial approximations which modelled QRS complex. This conclusion is valid for authentication and identification related with the approximations, excepted when the proposed system used as input the original records, regarding identification.

When comparing the work of [3] when using hermite approximations versus QRS complexes, it was observed that there was a performance decrease when using hermite approximations mainly with respect to the identification.

5.3.3 Cancelable Biometric System Proposed by [6]

The work of Dey et.al [6], described on section 3.2.1, was used in order to have a comparison term with a cancelable ECG approach. The methodology applied in the present work had some changes regarding the original method, which are explained also in section 3.2.1. Moreover, the algorithm used to detect P, R and P waves presented some flaws, resulting in a smallest number of subjects tested by this method, meaning this technique was applied to 155 subjects. The methodology developed by [6] is just applied to authentication, thus no identification performance indicators are showed in the presented section. To evaluate the system's performance, EER was computed regarding the set of tested discretization thresholds. Once more, for the sake of simplicity, in figure 5.12 just the EER average values are plotted.

Since the number of used subjects in this section was lower than the number regarding the tests related with the proposed methodology in the present work, it was decided that a new series of tests on the proposed system would be done using just that population (155 subjects), in order to obtain an actual comparison term. Once more, the tests were ran following the methodology explained on section 5.3.1 and by following **approach 1**, using EC combination. For the sake of simplicity, in figures just the EER average values are plotted.

The proposed system presented a better performance then the one proposed by [6] when using

Table 5.19: Computed **EER** average and standard deviation values regarding the method proposed by Dey et al. [6], when **hamming distance** is used as decision criterion.

Discretization Threshold	EER (%)
0.16	27.68 ± 0.85
0.47	13.40 ± 0.29
0.79	10.64 ± 0.46
1.11	9.90 ± 0.96
1.42	10.79 ± 0.79
1.74	16.96 ± 6.63
2.05	28.98 ± 1.35
2.37	39.88 ± 1.19
2.68	44.28 ± 1.26
3	47.36 ± 0.98

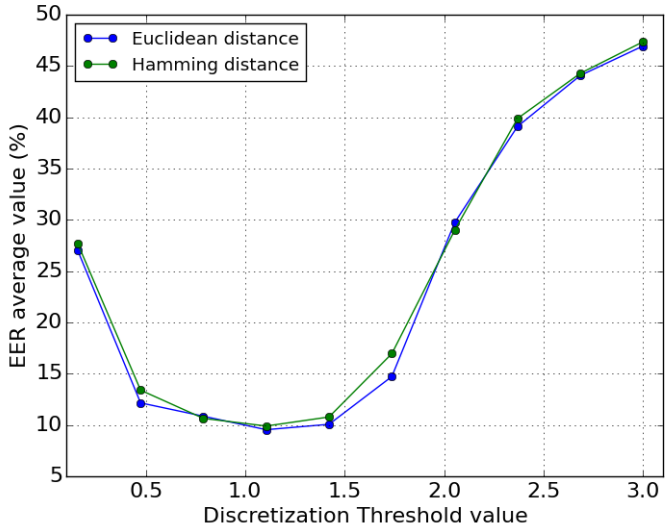

Figure 5.12: Computed **EER** average values regarding the method proposed by Dey et al. [6].

Table 5.20: Computed **EER** average and standard deviation values regarding population (155 subjects) used on method proposed by [6].

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	6.10 ± 0.13	7.95 ± 0.53	7.71 ± 0.71	7.11 ± 0.52	7.83 ± 1.00
20	5.67 ± 0.65	7.79 ± 0.52	7.46 ± 0.16	6.76 ± 0.22	7.01 ± 0.35
30	5.93 ± 0.31	7.21 ± 0.31	7.32 ± 0.25	6.58 ± 0.24	6.64 ± 0.17
40	5.73 ± 0.17	7.27 ± 0.09	7.25 ± 0.24	6.73 ± 0.10	6.76 ± 0.23
50	5.53 ± 0.34	7.27 ± 0.22	7.36 ± 0.13	6.58 ± 0.03	6.87 ± 0.31
60	5.62 ± 0.25	7.26 ± 0.32	7.21 ± 0.10	6.57 ± 0.06	6.81 ± 0.07
70	5.57 ± 0.29	7.17 ± 0.16	7.24 ± 0.18	6.59 ± 0.11	6.74 ± 0.17
80	5.75 ± 0.29	7.28 ± 0.30	7.24 ± 0.20	6.56 ± 0.03	6.67 ± 0.22
90	5.63 ± 0.32	7.22 ± 0.23	7.21 ± 0.16	6.58 ± 0.08	6.71 ± 0.29

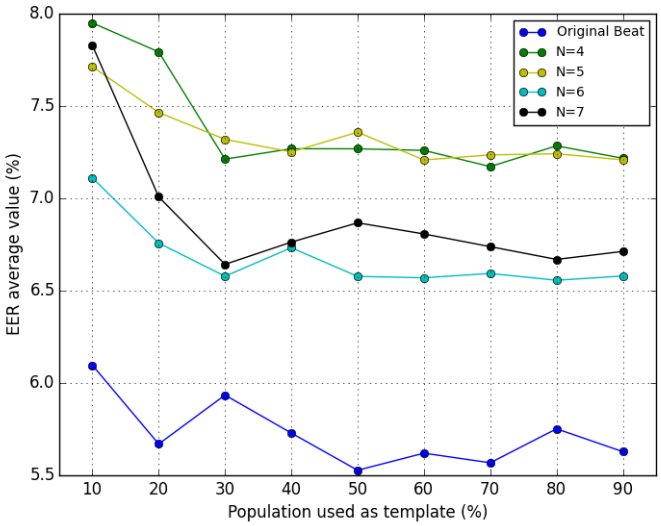

Figure 5.13: Computed **EER** and average values regarding population (155 subjects) used on method proposed by [6].

hamming distance as well as euclidean distance as metric. One must notice that the work of [6] does not mention which could be a good metric to use, since their final result was just to generate the cancelable templates, not applying the authentication process itself. However, it is important to take into account that the methodology used on the present work had some changes regarding the original one, which may introduce some decrease on the performance of the system.

5.4 Biometric Menagerie

As said in section 2.5, it can be very useful to observe the influence of the used database on the system one is developing. Thus, the biometric menagerie technique proposed by [29] was applied to HSM database. To do that the following steps were made:

1. The technique was directly applied to the results regarding the system proposed by Carreiras et.al [3].

2. The system developed by Carreiras et.al [3] and the proposed method were then applied to each species group.

This technique was applied following **approach 1**, and k set to 1, meaning 612 subjects were taken into consideration.

The division into different species and the renewed application of the proposed system produced the expected results, meaning: the subjects included in dove group were able to perform better, mainly when the system worked on authentication mode. However, the identification results were not very satisfactory. This fact can be justified by a certain degree of database inter-variability which can be proven by the differences regarding performance across the different species. Besides that, when considering the

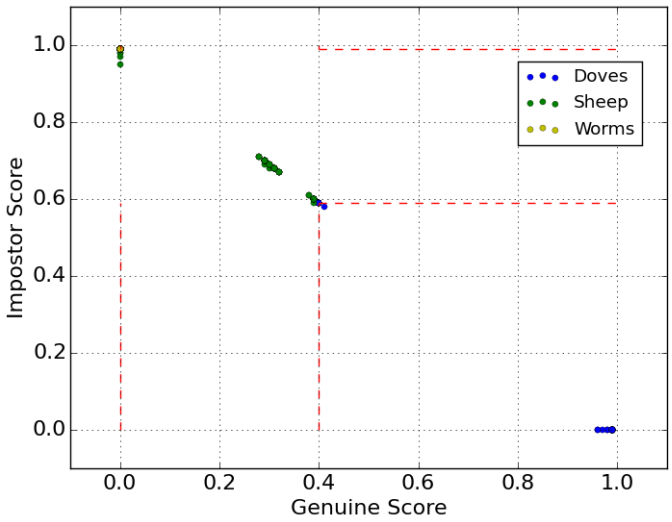

Figure 5.14: An example of the obtained zooplots. This plot results from the application of the biometric menagerie technique to the data regarding the modelling of the QRS complex, when six polynomials were used to perform it.

Table 5.21: Percentage average and standard deviation values of HSM database assigned to each species.

	Doves	**Sheep**	**Worms**	**Chameleons**	**Phantoms**
Original beat	73.40 $\pm$ 1.05	0	18.46 $\pm$ 1.09	0	8.14 $\pm$ 0.60
N=4	26.94 $\pm$ 1.81	12.44 $\pm$ 2.52	60.61 $\pm$ 1.28	0	0
N=5	26.96 $\pm$ 0.91	18.14 $\pm$ 0.94	54.90 $\pm$ 0.90	0	0
N=6	28.82 $\pm$ 1.49	19.82 $\pm$ 2.85	51.36 $\pm$ 1.82	0	0
N=7	29.01 $\pm$ 1.27	19.85 $\pm$ 1.86	51.13 $\pm$ 1.27	0	0

Table 5.22: Computed **EER** values arising from the renewed application of the system proposed by [3].

	Doves	**Sheep**	**Worms**	**Chameleons**	**Phantoms**
Original beat	3.67 $\pm$ 0.19	-	18.52 $\pm$ 1.76	-	6.60 $\pm$ 2.15
N=4	7.38 $\pm$ 0.59	12.50 $\pm$ 0.57	16.60 $\pm$ 0.24	-	-
N=5	6.50 $\pm$ 0.52	13.08 $\pm$ 0.44	16.47 $\pm$ 0.51	-	-
N=6	5.86 $\pm$ 0.47	12.81 $\pm$ 0.81	16.69 $\pm$ 1.02	-	-
N=7	6.82 $\pm$ 0.64	12.45 $\pm$ 1.55	16.58 $\pm$ 0.21	-	-

Table 5.23: Computed PC_{ID} values arising from the application of the system proposed by [3].

	Doves	**Sheep**	**Worms**	**Chameleons**	**Phantoms**
Original beat	4.34 $\pm$ 0.71	-	23.25 $\pm$ 1.99	-	3.41 $\pm$ 1.89
N=4	30.63 $\pm$ 0.29	31.03 $\pm$ 1.46	44.66 $\pm$ 0.57	-	-
N=5	27.33 $\pm$ 1.03	32.66 $\pm$ 1.04	43.68 $\pm$ 0.72	-	-
N=6	15.13 $\pm$ 2.72	36.71 $\pm$ 0.65	42.42 $\pm$ 0.38	-	-
N=7	27.90 $\pm$ 1.03	32.08 $\pm$ 1.53	42.35 $\pm$ 1.03	-	-

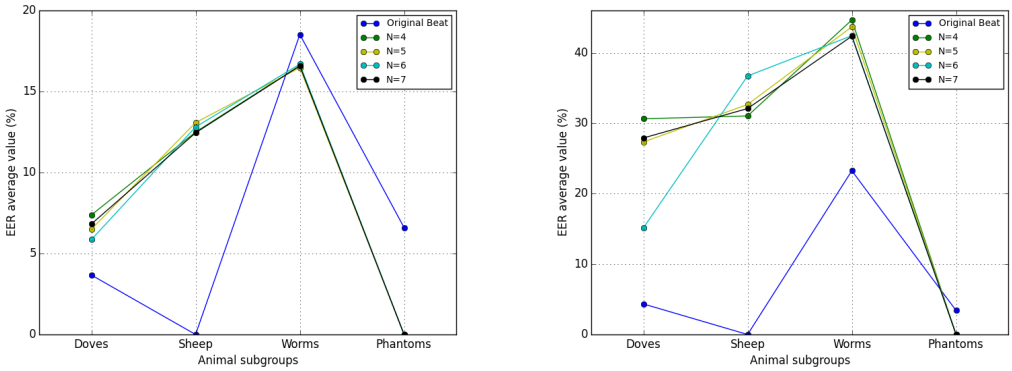

Figure 5.15: Computed **EER** and PC_{ID} average values regarding each animal subgroups, when a renewed application of the system proposed by [3] is applied.

Table 5.24: Computed **EER** average and standard deviation values regarding the renewed application of the proposed system, when **original heartbeats** are used as input to the proposed system.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	3.39 ± 0.25	-	7.87 ± 0.42	-	7.96 ± 1.15
20	3.30 ± 0.27	-	7.28 ± 0.70	-	6.02 ± 0.81
30	3.15 ± 0.25	-	6.88 ± 0.71	-	5.17 ± 0.64
40	3.28 ± 0.19	-	6.80 ± 0.71	-	5.47 ± 0.61
50	3.23 ± 0.15	-	7.06 ± 0.65	-	4.95 ± 0.67
60	3.15 ± 0.17	-	6.75 ± 0.42	-	4.72 ± 0.65
70	3.20 ± 0.14	-	6.73 ± 0.40	-	5.03 ± 0.85
80	3.18 ± 0.16	-	6.79 ± 0.48	-	4.87 ± 0.93
90	3.17 ± 0.18	-	6.76 ± 0.36	-	4.84 ± 0.80

Table 5.25: Computed PC_{ID} average and standard deviation values regarding the renewed application of the proposed system, when **original heartbeats** are used as input to the proposed system.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	7.64 ± 0.47	-	16.13 ± 1.07	-	17.35 ± 1.75
20	7.08 ± 0.41	-	13.50 ± 1.21	-	11.32 ± 2.08
30	7.05 ± 0.25	-	12.50 ± 0.33	-	9.89 ± 1.22
40	7.19 ± 0.16	-	12.37 ± 0.77	-	9.33 ± 0.92
50	7.09 ± 0.27	-	12.44 ± 0.38	-	9.09 ± 0.87
60	7.10 ± 0.27	-	11.67 ± 0.57	-	8.20 ± 0.71
70	6.96 ± 0.44	-	11.43 ± 0.77	-	7.90 ± 1.04
80	7.03 ± 0.27	-	11.62 ± 0.43	-	8.21 ± 0.76
90	7.03 ± 0.43	-	11.42 ± 0.64	-	8.21 ± 0.87

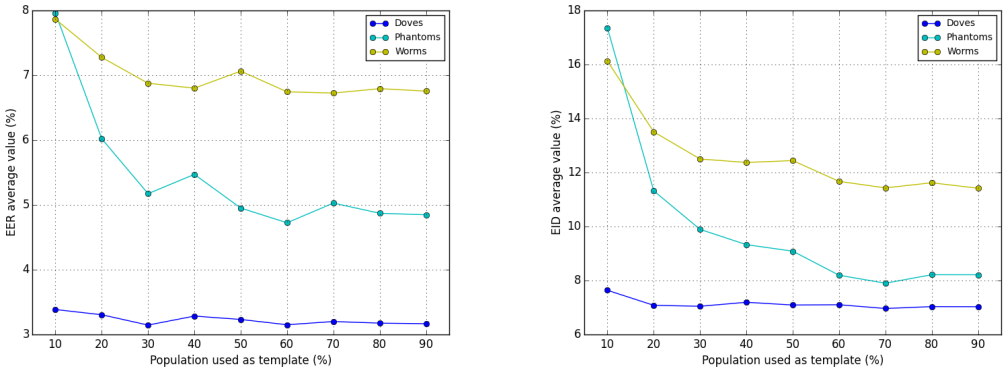

Figure 5.16: Computed **EER** and PC_{ID} average values regarding the renewed application of the proposed system, when **original heartbeats** are used as input to the proposed system.

Table 5.26: Computed **EER** average and standard deviation values regarding the renewed application of the proposed system, when **four polynomials** are used to model the QRS complex.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	4.63 ± 0.19	8.14 ± 1.40	8.32 ± 0.46	-	-
20	4.30 ± 0.31	6.18 ± 1.11	8.37 ± 0.36	-	-
30	4.17 ± 0.28	5.56 ± 1.04	8.38 ± 0.07	-	-
40	4.14 ± 0.26	5.50 ± 0.93	8.33 ± 0.31	-	-
50	4.06 ± 0.27	5.25 ± 0.54	8.38 ± 0.32	-	-
60	4.15 ± 0.26	5.28 ± 0.47	8.30 ± 0.26	-	-
70	4.03 ± 0.29	5.33 ± 0.56	8.26 ± 0.20	-	-
80	4.07 ± 0.23	5.18 ± 0.57	8.31 ± 0.25	-	-
90	4.04 ± 0.25	5.14 ± 0.52	8.30 ± 0.28	-	-

Table 5.27: Computed PC_{ID} average values regarding the renewed application of the proposed system, when **four polynomials** are used to model the QRS complex.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	15.33 ± 0.91	20.23 ± 1.16	38.25 ± 1.48	-	-
20	12.92 ± 1.74	19.53 ± 1.50	37.38 ± 0.49	-	-
30	12.02 ± 1.41	18.40 ± 0.88	37.42 ± 0.46	-	-
40	12.35 ± 1.30	18.27 ± 0.54	37.45 ± 0.25	-	-
50	11.12 ± 1.35	18.36 ± 0.22	37.46 ± 0.30	-	-
60	11.21 ± 0.79	18.48 ± 0.55	37.29 ± 0.30	-	-
70	11.15 ± 1.17	18.24 ± 0.19	37.48 ± 0.31	-	-
80	10.97 ± 0.91	18.28 ± 0.26	37.33 ± 0.34	-	-
90	10.91 ± 0.81	18.44 ± 0.31	37.32 ± 0.34	-	-

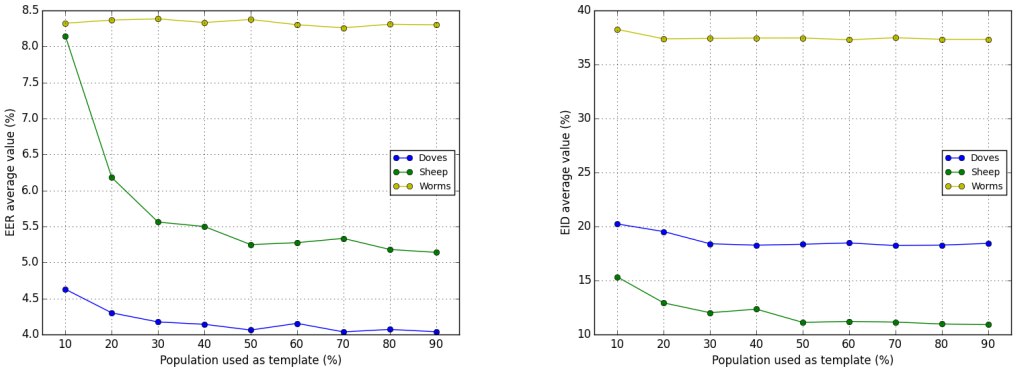

Figure 5.17: Computed **EER** and PC_{ID} average and standard deviation values regarding the renewed application of the proposed system, when **four polynomials** are used to model the QRS complex.

Table 5.28: Computed **EER** average and standard deviation values regarding the renewed application of the proposed system, when **five polynomials** are used to model the QRS complex.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	4.19 ± 0.29	6.74 ± 1.67	8.17 ± 0.29	-	-
20	4.11 ± 0.50	5.35 ± 0.21	8.21 ± 0.38	-	-
30	3.65 ± 0.41	5.17 ± 0.19	8.13 ± 0.09	-	-
40	3.81 ± 0.32	4.99 ± 0.16	8.00 ± 0.25	-	-
50	3.76 ± 0.35	4.98 ± 0.15	7.98 ± 0.10	-	-
60	3.82 ± 0.35	5.01 ± 0.10	7.97 ± 0.18	-	-
70	3.78 ± 0.23	5.01 ± 0.10	7.95 ± 0.12	-	-
80	3.70 ± 0.26	5.01 ± 0.11	7.95 ± 0.15	-	-
90	3.70 ± 0.31	4.99 ± 0.14	7.96 ± 0.12	-	-

Table 5.29: Computed PC_{ID} average and standard deviation values regarding the renewed application of the proposed system, when **five polynomials** are used to model the QRS complex.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	15.22 ± 0.61	20.06 ± 1.78	36.83 ± 0.38	-	-
20	13.79 ± 0.95	19.71 ± 1.05	36.55 ± 0.77	-	-
30	13.57 ± 0.47	18.67 ± 0.90	36.87 ± 0.30	-	-
40	14.11 ± 0.81	19.26 ± 0.46	36.83 ± 0.28	-	-
50	13.36 ± 0.89	18.19 ± 0.62	36.82 ± 0.19	-	-
60	13.54 ± 0.88	18.60 ± 0.38	36.79 ± 0.32	-	-
70	13.44 ± 0.56	18.39 ± 0.60	26.84 ± 0.19	-	-
80	13.37 ± 0.58	18.67 ± 0.20	36.79 ± 0.25	-	-
90	13.52 ± 0.65	18.75 ± 0.55	36.77 ± 0.24	-	-

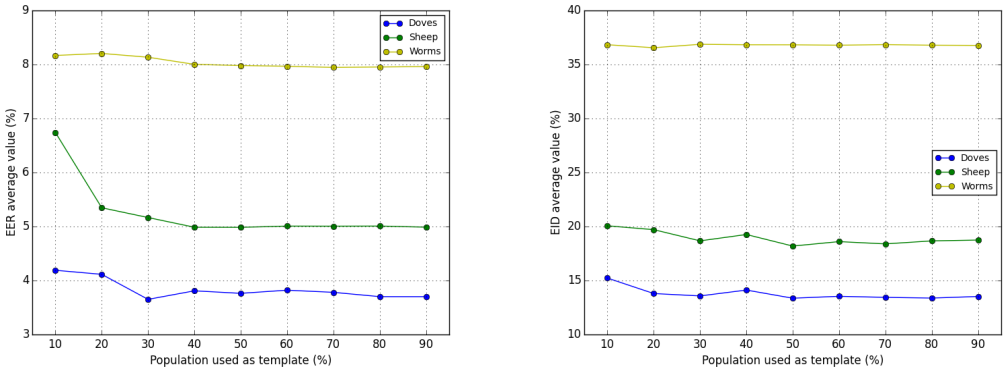

Figure 5.18: Computed **EER** and PC_{ID} average values regarding the renewed application of the proposed system, when **five polynomials** are used to model the QRS complex.

Table 5.30: Computed **EER** average and standard deviation values regarding the renewed application of the proposed system, when **six polynomials** are used to model the QRS complex.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	3.69 ± 1.04	6.66 ± 0.54	8.19 ± 0.32	-	-
20	3.19 ± 0.33	6.21 ± 0.49	8.18 ± 0.39	-	-
30	2.91 ± 0.49	6.26 ± 0.56	8.09 ± 0.49	-	-
40	2.84 ± 0.33	6.29 ± 0.74	8.17 ± 0.26	-	-
50	2.80 ± 0.32	6.22 ± 0.45	8.12 ± 0.32	-	-
60	2.77 ± 0.32	6.26 ± 0.59	8.11 ± 0.31	-	-
70	2.82 ± 0.34	6.18 ± 0.50	8.11 ± 0.40	-	-
80	2.85 ± 0.31	6.19 ± 0.52	8.05 ± 0.31	-	-
90	2.85 ± 0.30	6.16 ± 0.58	8.07 ± 0.31	-	-

Table 5.31: Computed PC_{ID} average and standard deviation values regarding the renewed application of the proposed system, when **six polynomials** are used to model the QRS complex.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	12.23 ± 0.90	27.27 ± 1.84	36.98 ± 0.41	-	-
20	11.28 ± 1.64	24.86 ± 0.82	36.45 ± 0.81	-	-
30	10.40 ± 0.61	25.26 ± 0.32	37.05 ± 0.34	-	-
40	10.45 ± 0.59	25.34 ± 0.45	36.43 ± 1.01	-	-
50	10.70 ± 0.60	25.51 ± 0.61	36.95 ± 0.22	-	-
60	10.32 ± 0.66	25.35 ± 0.58	36.98 ± 0.29	-	-
70	10.32 ± 0.97	25.48 ± 0.27	36.84 ± 0.23	-	-
80	10.21 ± 0.74	25.32 ± 0.59	36.99 ± 0.32	-	-
90	10.23 ± 0.86	25.39 ± 0.35	37.02 ± 0.13	-	-

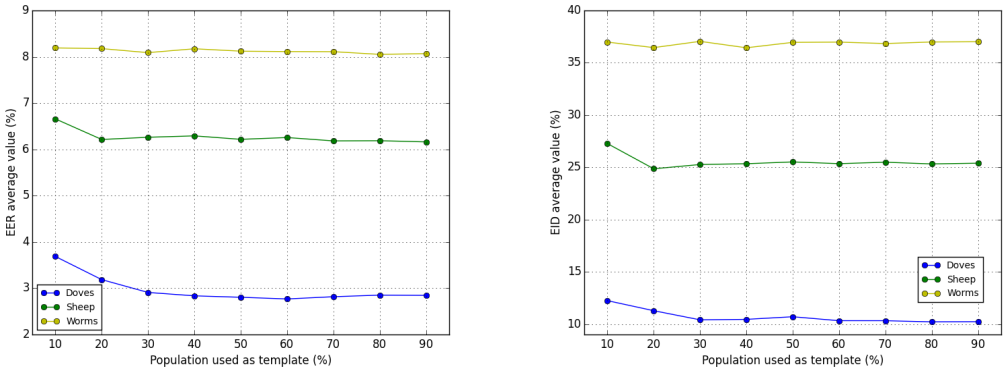

Figure 5.19: Computed **EER** and PC_{ID} average values regarding the renewed application of the proposed system, when **six polynomials** are used to model the QRS complex.

Table 5.32: Computed **EER** average and standard deviation values regarding the renewed application of the proposed system, when **seven polynomials** are used to model the QRS complex.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	4.54 ± 0.40	5.77 ± 0.51	8.11 ± 0.20	-	-
20	4.34 ± 0.27	5.08 ± 0.58	8.19 ± 0.21	-	-
30	4.15 ± 0.42	5.24 ± 0.48	8.40 ± 0.35	-	-
40	4.19 ± 0.41	5.05 ± 0.67	8.16 ± 0.25	-	-
50	4.27 ± 0.36	5.03 ± 0.60	8.16 ± 0.12	-	-
60	4.21 ± 0.36	5.01 ± 0.52	8.12 ± 0.24	-	-
70	4.23 ± 0.39	5.07 ± 0.51	8.22 ± 0.23	-	-
80	4.24 ± 0.35	4.97 ± 0.53	8.15 ± 0.23	-	-
90	4.22 ± 0.34	4.97 ± 0.53	8.17 ± 0.22	-	-

Table 5.33: Computed PC_{ID} average and standard deviation values regarding the renewed application of the proposed system, when **seven polynomials** are used to model the QRS complex.

per (%)	Doves	Sheep	Worms	Chameleons	Phantoms
10	16.67 ± 0.88	24.26 ± 1.66	36.59 ± 0.52	-	-
20	15.47 ± 0.96	22.34 ± 0.76	35.73 ± 1.02	-	-
30	14.98 ± 0.44	23.00 ± 0.65	36.27 ± 0.76	-	-
40	15.18 ± 0.64	22.31 ± 0.79	35.71 ± 1.10	-	-
50	15.24 ± 0.25	22.06 ± 1.30	35.72 ± 1.14	-	-
60	15.29 ± 0.32	22.36 ± 0.71	35.58 ± 1.16	-	-
70	15.13 ± 0.40	21.65 ± 0.39	35.68 ± 1.13	-	-
80	15.14 ± 0.35	22.01 ± 1.06	35.73 ± 1.03	-	-
90	15.14 ± 0.43	22.39 ± 1.29	36.69 ± 0.14	-	-

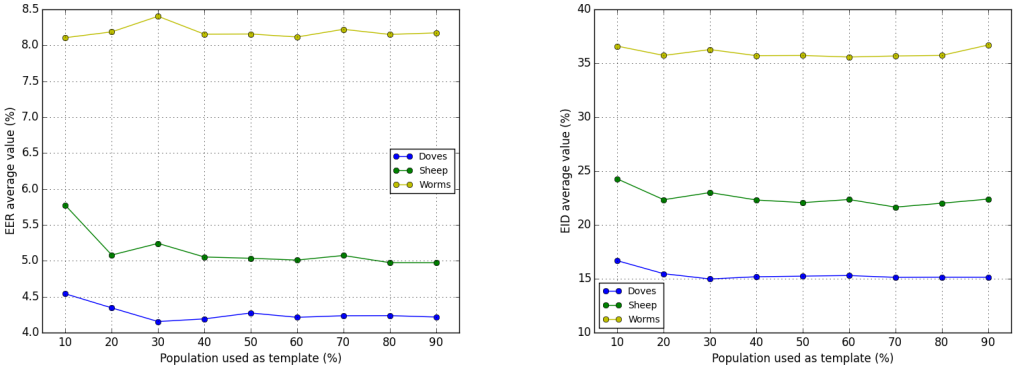

Figure 5.20: Computed **EER** and PC_{ID} average values regarding the renewed application of the proposed system, when **seven polynomials** are used to model the QRS complex.

approximations which modelled the QRS complex, the technique assigned a considerably percentage of the database to the worm group, which by definition, it is the worst group to belong to.

Another important aspect to notice is that system's renewed application allowed to observe a more established tendency regarding per and the number of hermite polynomials used to modelling the QRS complex. Meaning, it is fair to say that with the increase of both per and number of polynomials used, the recognition results (mainly regarding authentication) were better. This was particularly true in dove and sheep subgroups.

Chapter 6

Conclusions

Nowadays, security and robustness are two of the main key factors regarding the application of biometrics systems in real life situations. Therefore it is crucial to develop systems able to convey confidence to the users, maintaining a good recognition performance. The present work aimed to develop a cancelable biometric system. To do that, two main fields were approached. On the one hand, hermite polynomial approximations modelling QRS complex and one the other hand, the dissimilarity based representation. These two fields are able, in separated or set together, to ensure a cancelable biometric system, namely by guaranteeing *irreversibility* and *unlinkability*.

Since the database had already filtered signals, this process was out of the context of this work. Besides that, one of the stages of the proposed system was ECG segmentation into heartbeats. There are several algorithms to perform this task, thus in order to save time, it was chosen to apply an already developed algorithm proposed by [11]. Note that the obtained results are obviously related with the adopted methodology but also with the database used to apply it.

Regarding hermite polynomial approximations, used to modelling the QRS complex, it was possible to conclude that using it as input to the proposed system conducted to a slightly decrease in recognition performance, when compared with the system working with the original records. Besides that, the influence of the number of used polynomials was studied. It was not completely clear that using a higher number of polynomials conducted to a better system's performance.

Regarding dissimilarity based representation and classifier, besides the influence of the dissimilarity measure and of the classification metric, the influence of the number of neighbours, k, used by the k-NN and the percentage of population used as templates, per, was studied. Besides studying the influence of the parameters referred above, also the way of how database records were used was studied, as depicted in section 5.3.1.

It was possible to conclude that the proposed method, globally performed better than the other mentioned biometric systems when using the original records as well as when using the hermite polynomial approximations. Note that all the tests were performed using the same dataset. It is a very important result, since besides performing better than a traditional biometric system like the one proposed by [3], it ensures the protection of the biometric data, which is a crucial point regarding the security needed in

the present recognition systems, while maintaining or even improving the recognition performance.

The application of the biometric menagerie aimed to study the database quality with respect to its subjects. By analysing the obtained results, it was possible to conclude that most of the population belonged to the dove subgroup, when the original records were considered. Regarding the hermite polynomial approximations, approximately half of the population was assigned to the worm subgroup which can explain the slightly decrease in the system's performance related with authentication and the not so good results regarding the identification process.

6.1 Future Work

To implement a system of this kind in a real life application is a long way. The present work just pointed out a possible solution in a very early stage of development.

It is very important to perform these kind of tests in situations close to the real one. Meaning, having a larger database where each subject has several records will give a more concrete idea if the proposed system has in fact conditions to be applied in a real life situation.

As depicted above, one of the main issues regarding biometrics applied to ECG is related with the intra-variability. One way to solve this question is, as said above, to have several records from the same individual. Thus, it is important to take into account the moment of the day when the records are being produced or, for example, the subject's emotional state.

Other of the themes that can be explored is the quality of the hermite approximation, meaning that it would be useful to apply a different algorithm in order to minimize equation 2.12. With more available time and a powerful system, the use of more hermite polynomials to perform the approximation will be useful too.

Regarding the dissimilarity based classifier, the use of other biometric features as input, such as fingerprint or iris will allow to conclude its effectiveness, since those two features are in a more mature state regarding the application of cancelable biometrics.

References

[1] E. C. Bank. Third report on card fraud. Technical report, European Central Bank, 2014.

[2] L. Biel, O. Pettersson, L. Philipson, and P. Wide. Ecg analysis: a new approach in human identification. *Instrumentation and Measurement, IEEE Transactions on*, 50(3):808–812, 2001.

[3] C. Carreiras, A. Lourenco, A. Fred, and R. Ferreira. Ecg signals for biometric applications-are we there yet? In *Informatics in Control, Automation and Robotics (ICINCO), 2014 11th International Conference on*, volume 2, pages 765–772. IEEE, 2014.

[4] R. Clarke. Human identification in information systems: Management challenges and public policy issues. *Information Technology & People*, 7(4):6–37, 1994.

[5] D. P. Coutinho, H. Silva, H. Gamboa, A. Fred, and M. Figueiredo. Novel fiducial and non-fiducial approaches to electrocardiogram-based biometric systems. *IET biometrics*, 2(2):64–75, 2013.

[6] N. Dey, B. Nandi, M. Dey, D. Biswas, A. Das, and S. S. Chaudhuri. Biohash code generation from electrocardiogram features. In *Advance Computing Conference (IACC), 2013 IEEE 3rd International*, pages 732–735. IEEE, 2013.

[7] R. P. Duin and E. Pekalska. The dissimilarity representation for structural pattern recognition. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 1–24. Springer, 2011.

[8] R. P. Duin and E. Pekalska. The dissimilarity space: Bridging structural and statistical pattern recognition. *Pattern Recognition Letters*, 33(7):826–832, 2012.

[9] M. Francisco, C. Carlos, Lourenço, F. Ana, and F. Rui. ECG Biometrics Using a Dissimilarity Space Representation. 2015.

[10] J. E. Hall. *Guyton and Hall textbook of medical physiology*. Elsevier Health Sciences, 2010.

[11] P. Hamilton. Open source ecg analysis. In *Computers in Cardiology, 2002*, pages 101–104. IEEE, 2002.

[12] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold. Ecg to identify individuals. *Pattern recognition*, 38(1):133–142, 2005.

[13] A. Jain, R. Bolle, and S. Pankanti. *Biometrics: personal identification in networked society*, chapter Introduction to Biometrics, pages 1–41. Springer Science & Business Media, 2006.

[14] A. K. Jain and A. Kumar. Biometrics of next generation: An overview. *Second Generation Biometrics*, 12(1):2–3, 2010.

[15] A. K. Jain, K. Nandakumar, and A. Nagar. Biometric template security. *EURASIP Journal on Advances in Signal Processing*, 2008:113, 2008.

[16] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo. Clustering ecg complexes using hermite functions and self-organizing maps. *Biomedical Engineering, IEEE Transactions on*, 47(7):838–848, 2000.

[17] E. Maiorana, P. Campisi, J. Fierrez, J. Ortega-Garcia, and A. Neri. Cancelable templates for sequence-based biometrics with application to on-line signature recognition. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(3):525–538, 2010.

[18] D. G. Márquez, A. Otero, P. Félix, and C. a. García. On the Accuracy of Representing Heartbeats with Hermite Basis Functions. *Biosignals*, pages 338–341, 2013.

[19] A. Nagar and A. K. Jain. On the security of non-invertible fingerprint template transforms. In *Information Forensics and Security, 2009. WIFS 2009. First IEEE International Workshop on*, pages 81–85. IEEE, 2009.

[20] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. O'Sullivan, E. J. Sirevaag, S. D. Kristjansson, A. K. Sheffield, J. W. Rohrbaugh, et al. Ecg biometrics: A robust short-time frequency analysis. In *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*, pages 1–6. IEEE, 2010.

[21] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh. Ecg biometric recognition: A comparative analysis. *Information Forensics and Security, IEEE Transactions on*, 7(6):1812–1824, 2012.

[22] O. Ouda, N. Tsumura, and T. Nakaguchi. Tokenless cancelable biometrics scheme for protecting iris codes. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 882–885. IEEE, 2010.

[23] S. Pathoumvanh, S. Airphaiboon, and K. Hamamoto. Robustness study of ecg biometric identification in heart rate variability conditions. *IEEJ Transactions on Electrical and Electronic Engineering*, 9(3):294–301, 2014.

[24] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*, 40(3):614–634, 2001.

[25] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle. Generating cancelable fingerprint templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):561–572, 2007.

[26] C. Rathgeb and A. Uhl. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP Journal on Information Security*, 2011(1):1–25, 2011.

[27] A. B. Teoh, Y. W. Kuan, and S. Lee. Cancellable biometrics and annotations on biohash. *Pattern recognition*, 41(6):2034–2044, 2008.

[28] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis. Analysis of human electrocardiogram for biometric recognition. *EURASIP journal on Advances in Signal Processing*, 2008:19, 2008.

[29] N. Yager and T. Dunstone. The biometric menagerie. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(2):220–230, 2010.

Appendix A

Histogram Superposition Areas Tables

Table A.1: Average and standard deviation superposition area values regarding **CC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	0.077 ± 0.021	0.188 ± 0.136	0.066 ± 0.014	0.078 ± 0.020	0.142 ± 0.010
20	0.071 ± 0.016	0.084 ± 0.059	0.076 ± 0.017	0.072 ± 0.016	0.123 ± 0.064
30	0.078 ± 0.013	0.114 ± 0.073	0.086 ± 0.011	0.085 ± 0.005	0.121 ± 0.055
40	0.082 ± 0.001	0.139 ± 0.055	0.079 ± 0.015	0.081 ± 0.006	0.134 ± 0.032
50	0.079 ± 0.008	0.140 ± 0.051	0.086 ± 0.004	0.081 ± 0.010	0.111 ± 0.022
60	0.080 ± 0.008	0.136 ± 0.045	0.086 ± 0.005	0.084 ± 0.005	0.112 ± 0.027
70	0.083 ± 0.007	0.148 ± 0.037	0.083 ± 0.005	0.084 ± 0.006	0.122 ± 0.024
80	0.082 ± 0.006	0.015 ± 0.016	0.088 ± 0.004	0.087 ± 0.001	0.119 ± 0.018
90	0.081 ± 0.006	0.016 ± 0.011	0.087 ± 0.005	0.086 ± 0.003	0.116 ± 0.015

Table A.2: Average and standard deviation superposition area values regarding **CE** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	0.106 ± 0.004	0.088 ± 0.004	0.095 ± 0.003	0.093 ± 0.003	0.091 ± 0.004
20	0.107 ± 0.004	0.088 ± 0.003	0.093 ± 0.002	0.092 ± 0.002	0.093 ± 0.002
30	0.108 ± 0.004	0.088 ± 0.003	0.095 ± 0.002	0.094 ± 0.002	0.094 ± 0.002
40	0.110 ± 0.004	0.089 ± 0.002	0.096 ± 0.003	0.094 ± 0.002	0.095 ± 0.002
50	0.112 ± 0.004	0.090 ± 0.002	0.096 ± 0.002	0.096 ± 0.003	0.095 ± 0.002
60	0.113 ± 0.004	0.091 ± 0.002	0.097 ± 0.002	0.097 ± 0.002	0.096 ± 0.002
70	0.115 ± 0.004	0.091 ± 0.002	0.099 ± 0.002	0.098 ± 0.002	0.098 ± 0.002
80	0.117 ± 0.004	0.092 ± 0.002	0.099 ± 0.003	0.099 ± 0.002	0.099 ± 0.002
90	0.117 ± 0.005	0.093 ± 0.002	0.100 ± 0.002	0.100 ± 0.002	0.099 ± 0.002

Table A.3: Average and standard deviation superposition area values regarding **EC** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	0.032 ± 0.004	0.046 ± 0.006	0.047 ± 0.004	0.043 ± 0.004	0.041 ± 0.003
20	0.030 ± 0.002	0.046 ± 0.003	0.044 ± 0.003	0.043 ± 0.001	0.043 ± 0.003
30	0.031 ± 0.003	0.045 ± 0.004	0.044 ± 0.003	0.042 ± 0.002	0.042 ± 0.002
40	0.031 ± 0.001	0.046 ± 0.002	0.043 ± 0.002	0.042 ± 0.002	0.042 ± 0.002
50	0.031 ± 0.002	0.046 ± 0.002	0.043 ± 0.001	0.041 ± 0.001	0.042 ± 0.001
60	0.031 ± 0.002	0.045 ± 0.001	0.044 ± 0.002	0.042 ± 0.001	0.043 ± 0.001
70	0.030 ± 0.002	0.045 ± 0.001	0.044 ± 0.001	0.043 ± 0.002	0.041 ± 0.001
80	0.031 ± 0.001	0.045 ± 0.001	0.044 ± 0.001	0.042 ± 0.001	0.042 ± 0.001
90	0.031 ± 0.001	0.045 ± 0.001	0.044 ± 0.001	0.043 ± 0.001	0.042 ± 0.001

Table A.4: Average and standard deviation superposition area values regarding **EE** combination.

per (%)	Original Beat	N=4	N=5	N=6	N=7
10	0.033 ± 0.003	0.028 ± 0.002	0.029 ± 0.002	0.029 ± 0.001	0.029 ± 0.002
20	0.036 ± 0.003	0.032 ± 0.002	0.032 ± 0.001	0.034 ± 0.002	0.032 ± 0.002
30	0.037 ± 0.003	0.036 ± 0.002	0.035 ± 0.002	0.036 ± 0.002	0.035 ± 0.001
40	0.038 ± 0.003	0.038 ± 0.002	0.038 ± 0.001	0.039 ± 0.002	0.038 ± 0.001
50	0.040 ± 0.002	0.041 ± 0.001	0.041 ± 0.002	0.042 ± 0.002	0.040 ± 0.001
60	0.042 ± 0.002	0.043 ± 0.002	0.042 ± 0.002	0.044 ± 0.002	0.043 ± 0.001
70	0.044 ± 0.003	0.045 ± 0.002	0.044 ± 0.002	0.046 ± 0.002	0.045 ± 0.001
80	0.045 ± 0.003	0.047 ± 0.002	0.047 ± 0.002	0.048 ± 0.002	0.046 ± 0.001
90	0.046 ± 0.002	0.049 ± 0.002	0.048 ± 0.002	0.049 ± 0.002	0.048 ± 0.001