ECG in Biometric Recognition: Time Dependency and Application Challenges

by

Foteini Agrafioti

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

Abstract

ECG in Biometric Recognition:

Time Dependency and Application Challenges

Foteini Agrafioti

Doctor of Philosophy

Graduate Department of Electrical and Computer Engineering

University of Toronto

2011

As biometric recognition becomes increasingly popular, the fear of circumvention, obfuscation and replay attacks is a rising concern. Traditional biometric modalities such as the face, the fingerprint or the iris are vulnerable to such attacks, which defeats the purpose of biometric recognition, namely to employ physiological characteristics for secure identity recognition.

This thesis advocates the use the electrocardiogram (ECG) signal for human identity recognition. The ECG is a vital signal of the human body, and as such, it naturally provides liveness detection, robustness to attacks, universality and permanence. In addition, ECG inherently satisfies uniqueness requirements, because the morphology of the signal is highly dependent on the particular anatomical and geometrical characteristics of the myocardium in the heart.

However, the ECG is a continuous signal, and this presents a great challenge to biometric recognition. With this modality, instantaneous variability is expected even within recordings of the same individual due to a variety of factors, including recording noise, or physical and psychological activity. While the noise and heart rate variations due to physical exercise can be addressed with appropriate feature extraction, the effects of emotional activity on the ECG signal are more obscure.

This thesis deals with this problem from an affective computing point of view. First,

the psychological conditions that affect the ECG and endanger biometric accuracy are identified. Experimental setups that are targeted to provoke *active* and *passive arousal* as well as *positive* and *negative valence* are presented. The empirical mode decomposition (EMD) is used as the basis for the detection of emotional patterns, after adapting the algorithm to the particular needs of the ECG signal. Instantaneous frequency and oscillation features are used for state classification in various clustering setups. The result of this analysis is the designation of psychological states which affect the ECG signal to an extent that biometric matching may not be feasible. An updating methodology is proposed to address this problem, wherein the signal is monitored for instantaneous changes that require the design of a new template.

Furthermore, this thesis presents the enhanced Autocorrelation- Linear Discriminant Analysis (AC/LDA) algorithm for feature extraction, which incorporates a signal quality assessment module based on the periodicity transform. Three deployment scenarios are considered namely a) small-scale recognition systems, b) large-scale recognition systems and c) recognition in distributed systems. The enhanced AC/LDA algorithm is adapted to each setting, and the advantages and disadvantages of each scenario are discussed.

Overall, this thesis attempts to provide the necessary algorithmic and practical framework for the real-life deployment of the ECG signal in biometric recognition.

Acknowledgements

First and foremost, I would like to sincerely thank my advisor Prof. Dimitrios Hatzinakos. He gave me guidance and direction that was needed to produce this work. Without his generous help, this research would not have been possible. I would also like to thank the Department of Electrical and Computer Engineering for their financial support.

Thank you to my proposal and defence committees for taking the time to provide useful insight. I am also grateful to the Communications Group faculty members for teaching, offering technical advice and inspiration in the beginning of this thesis work. Financial support for this thesis was provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Last but not least, I would like to thank my family and friends for their constant encouragement during my studies.

Contents

List of Acronyms

AC	Autocorrelation
ADM	Affective Dimensional Model
ANS	Autonomic Nervous System
AV	Arousal-Valence
BVP	Blood Volume Pressure
CRF	Conditional Random Field
DBNN	Decision Based Neural-Network
DCT	Discrete Cosine Transform
DEM	Discrete Emotional Model
DFA	Discriminant Function Analysis
DTW	Dynamic Time Warping
ECG	Electrocardiogram
EEG	Electroencephalogram
EER	Equal Error Rate
EMD	Empirical Mode Decomposition
EMG	Electromyogram
FAR	False Acceptance Rate
FLDA	Fisher Linear Discriminant Analysis
FP	Fisher Projection

FRR	False Rejection Rate
GMM	Gaussian Mixture Model
GSR	Galvanic Skin Response
HMM	Hidden Markov Model
HR	Heart-Rate
HRV	Heart-Rate Variability
IADS	International Affective Digitized Sounds
IAPS	International Affective Picture System
ICA	Independent Component Analysis
LDA	Linear Discriminant Analysis
MAUI	Multimodal Affect User Interface
MBP	Marquardt Backpropagation
MLP	Multilayer Perceptron
PCA	Principal Component Analysis
PPG	Phonocardiogram
PT	Periodicity Transform
SAM	Self-Assessment Manikin
SFFS	Sequential Floating Forward Search
SIMCA	Soft Independent Modelling of Class Analogy
SVM	Support Vector Machine

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction to Biometric Recognition

Automatic and accurate identity validation is becoming increasingly critical in several aspects of our every day lives, such as in financial transactions, access control, traveling, healthcare and many others. Traditional strategies for automatic identity recognition include items such as PIN numbers, tokens, passwords and ID cards. Despite the wide deployment of such tactics, the means for authentication is either *entity-based* or *knowledge-based* which raises serious concerns with regard to the risk of identity theft.

According to the latest US Federal Trade Commission report [6], in 2009 identity theft was the number one complaint category (a total of 721,418 cases of consumer complaints). Identity theft can take different forms - credit card fraud (17%), falsification of government documents (16%), utilities fraud (15%), employment fraud (13%) and others. Among these cases, true-identity theft constitutes only a small portion of the complaints, while ID falsification appears to be the greatest threat. Unfortunately, the technology for forgery advances without analogous improvements on the security side.

Biometric recognition was introduced as a more secure means of identity establishment. Biometric modalities are characteristics of the human body that are unique for

every individual and that can be used to establish his/her identity in a population. These characteristics can be either physiological or behavioral. For instance, the face, the iris and the fingerprints are physiological biometric modalities. Keystroke dynamics, the gait and the voice are examples of behavioral biometric modalities. The fact that biometric modalities are directly linked with the users presents an extraordinary opportunity to bridge the security gaps caused by traditional recognition strategies. Biometric modalities are difficult to steal or counterfeit when compared to PIN numbers or passwords. In addition, the convenience of not having to carry a piece of ID or remember a password makes biometric systems more accessible and easy to use.

Each biometric modality has unique characteristics. For instance, face pictures may be acquired from distance which makes them suitable for surveillance. On the other hand, fingerprints need direct contact with the sensing device. Every biometric feature has its own strengths and weaknesses and deployment choices are based on the characteristics of the envisioned application environment.

There is a major drawback with biometric recognition - as opposed to static PIN numbers or passwords, biometric recognition may present *false rejection* since usually no two readings of the same biometric modality are identical. Anatomical, psychological or even environmental factors affect the appearance of the biometric modality at any particular instance. For instance, faces may be presented to the recognizers under various expressions, different lighting settings or with occlusion (glasses, hats etc). This may introduce significant variability (commonly referred to as *intra-subject* or *intra-class* variability), and the challenge is to design algorithms that are robust to it.

Provided intra-subject variability can be addressed with appropriate feature extraction, another important consideration is the robustness to circumvention and replay attacks. Circumvention is a form of biometric forgery - for example, falsified fingerprints that are reproduced from an original fingerprint. A replay attack is the presentation to the system of the original biometric feature from an illegitimate subject, for example

pre-recorded voice playbacks in speaker recognition systems. Biometric obfuscation is another prominent risk with this technology. There are cases where biometric features are intentionally removed to avoid establishment of the true identity (for example asylum-seekers in Europe [7] removed their fingerprints to avoid identification). With the wide deployment of biometrics, these attacks are becoming frequent and concerns are being raised on the security levels that this technology can offer.

Concentrated efforts have been made for the development the next generation of biometric characteristics that are inherently robust to the above mentioned attacks. Characteristics that are internal to the human body have been investigated such as vein patterns, the odour and cognitive biometrics. The *medical biometrics* constitutes another category of new biometric modalities that encompasses signals which are typically used in clinical diagnostics. Some examples of medical biometric signals are the electrocardiogram (ECG), phonocardiogram (PPG), electroencephalogram (EEG), blood volume pressure (BVP) and electromyogram (EMG).

Medical biometrics have been actively investigated only within the last decade. Although the biometric property of these signals had been observed before, the complicated signal acquisition process and the waiting times were restrictive for application in access control. However, with the development of dry recoding sensors that are easy to attach even by non-trained personnel, the medical biometrics field flourished. The rapid advancement over the last decade was supported by the fact that signal-processing tools had already been developed for diagnostic purposes.

The main advantage of medical biometrics is the robustness to circumvention, replay and obfuscation attacks. If established as biometrics, then the respective systems are empowered with an inherent shield to such threats. Another advantage of medical biometrics is the possibility of utilizing them for continuous authentication, since they can provide a fresh biometric reading every couple of seconds. In addition, medical biometrics are one dimensional physiological signals, which ensures low computational effort.

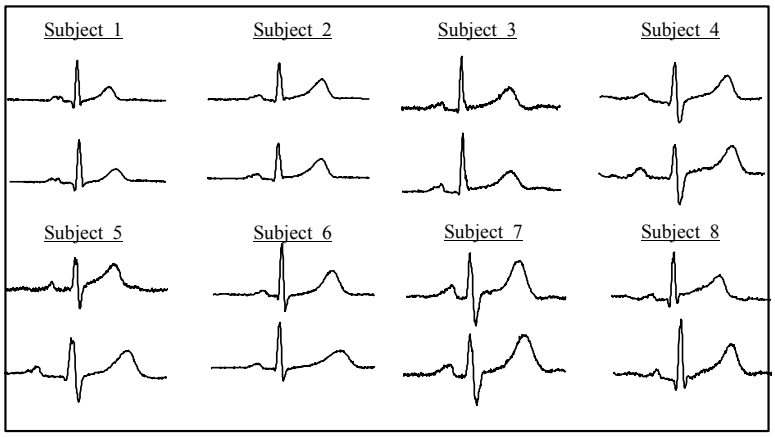


Figure 1.1: Heart Beats of the same individual recorded a few years apart. Heart beats have been drawn from the PTB diagnostic database [4]

In this work, we study the ECG signal. The concepts presented herein, however, may be extended to all medical biometric modalities.

1.2 ECG Biometrics: Motivation and Challenges

The ECG signal describes the variation of electrical activity of the heart over time. It is recorded non-invasively with electrodes attached at the surface of the body. Traditionally, physicians use the ECG to gain insight on heart conditions, while usually complementary tests are required to finalize a diagnosis. However, from a biometrics perspective, it has been demonstrated that the ECG has sufficient detail for identification.

Among the desirable properties of the ECG biometric modality are universality, permanence, uniqueness, robustness to attacks, liveness detection, continuous authentication and data minimization. More precisely,

1. *Universality* refers to the ability of collecting the biometric sample from the general population. Since the ECG is a vital signal, this property is satisfied naturally.

2. *Permanence* refers to the ability of performing biometric matches against templates

that have been designed earlier in time. This essentially requires that the signal is stable over time. Figure 1.1 shows an example of signal stability from the PTB database [4], which offers signals for the same individual collected a few years apart. As will be discussed later, the ECG is affected by both physical and psychological activity. Even though the specific local characteristics of the pulses may change, the overall diacritical waves and morphologies are still observable.

3. *Uniqueness* is guaranteed in the ECG signal because of its physiological origin. While ECG signals of different individuals conform to approximately the same pattern, there is large inter-individual variability due to the various electrophysiological parameters that control the generation of this waveform.

4. *Robustness to attacks.* The particular appearance of the ECG waveform is the outcome of several sympathetic and parasympathetic factors of the human body. Controlling the waveform or attempting to mimic somebody else's ECG signal is extremely difficult, if not impossible. To the best of our knowledge, there is currently no means of falsifying an ECG waveform and presenting it to a biometric recognition system. Obfuscation is also addressed naturally.

5. *Liveness detection.* ECG offers natural liveness detection, being only present in a *living* subject. With this modality the recognizer can trivially ensure sensor liveness. Other biometric modalities, such the iris or the fingerprint require additional processing to establish the liveness of the reading.

6. *Continuous authentication.* As opposed to static iris or fingerprint images, the ECG is a dynamic biometric modality that evolves with time. When deployed for security in welfare monitoring environments, a fresh reading can be obtained every couple of seconds to re-authenticate an identity. This property is unique to medical biometrics.

7. *Data minimization.* Privacy intrusion is becoming increasingly critical in environments of airtight security. One way to address this problem is to utilize as few identifying credentials as possible. Data minimization is possible with ECG biometrics because there are environments where the collection of the signal is performed irrespective of the identification task. Examples of such environments are tele-medicine, patient monitoring in hospitals, field agent monitoring (fire-fighters, policemen, soldiers etc).

Despite the advantages, notable challenges arise with this technology when large-scale deployment is envisioned:

1. *Time dependency.* With time-varying biosignals there is high risk of instantaneous changes which may endanger biometric security. Recordings of the cardiac potential at the surface of the body are very prone to noise due to body movements. However, even in the absence of noise, the ECG signal may destabilize with respect to a biometric template that was constructed some time earlier. The reason for this is the direct effect that the body's physiology and psychology have on the cardiac function. Therefore, a central aspect of the ECG biometrics research is the investigation of the sources of intra-subject variability.

2. *Collection periods.* As opposed to biometrics such as the face, the iris or the fingerprint, where the biometric information is available for capturing at any time instance, this is not the case with the ECG signal. Every heart beat is formed within approximately a second, which essentially means that longer waiting times are expected with this technology, especially when long ECG segments are required for feature extraction. The challenge is to minimize the number of pulses that an algorithm uses for recognition, as well as the processing time.

3. *Privacy implications.* When collecting ECG signals a large amount of sensitive information is inevitably collected. The ECG signal may reveal current and past

medical conditions, as well as hints about the emotional state of the monitored individual. Traditionally, the ECG is available to physicians only. Thus, the possibility of linking ECG samples to identities raises serious privacy issues.

4. *Cardiac Conditions.* Although cardiac disorders are not as frequent a damaging factor as injuries for more conventional biometrics (fingerprint, face), they can limit ECG biometric methods. Disorders can range from an isolated irregularity (atria and ventricle premature contractions) to severe conditions which require immediate medical assistance. The biometric challenge is therefore to design algorithms that are invariant to everyday ECG irregularities [8].

1.3 Research Goals and Contributions

There are a number of technical challenges that have motivated the work presented in this thesis. Overall, the necessary framework for real life deployment of ECG biometric recognition is provided from an algorithmic and implementation point of view. While all application possibilities are discussed, the main interest of this work is in securing welfare monitoring environments, where the user's identity is authenticated continuously. It is anticipated that security within such settings will be one the most prominent application of this technology. The contributions of the present work can be summarized as follows:

- Design of an efficient algorithm for ECG-based recognition. Our prior work in this field [9] was the basis for the development of signal quality assessment methodologies, which preprocess the ECG signal in-hand before biometric matching. In addition, the original AC/LDA algorithm has been improved to further address physiological variations of the signal namely the Heart Rate Variability (HRV). The enhanced algorithm is evaluated using databases of ECG signals that were collected at the Biometrics Security Laboratory, at the University of Toronto.

- The application frameworks that this technology can fit in are defined and the respective technical challenges are addressed. Three distinct application environments are identified namely, A) small scale access control, B) large scale recognition and C) recognition in distributed systems. The recognition algorithm has been adjusted to every setting.

- The third contribution of this thesis is the identification of psychological factors that may compromise an ECG biometric template. In welfare monitoring, the time-dependent property of the ECG biometric is two-fold. While a new reading can be collected and used for continuous authentication, emotional factors may destabilize the signal. A first step of this analysis was to demonstrate the perils of ignoring time-dependency. Experiments that simulate real life monitoring environments were performed, and it was observed that in the absence of noise and physical activity, the waveform of the ECG signal may still vary due to the effects of the autonomic nervous system (ANS).

- The above issue is extensively investigated in this thesis by examining the feasibility of detecting human emotion from the ECG signal (*affective computing*). A new approach to emotional pattern recognition is proposed based on the Empirical Mode Decomposition (EMD). The decomposition is first refined for the ECG case, while the analysis is performed on signals that were collected by inducing both active and passive arousal, as well as valence (see Appendix B for definitions). This study indicates that ECG emotional variation is subject-specific as well as most prominent under active arousal. These findings are of great importance for real life deployment of ECG biometrics.

- A method to automatically detect the destabilization of the ECG template is also proposed in this work, by incorporating the above findings into the recognition framework. It is demonstrated that the detected emotional states correspond to

portions the signal characterized by sufficient biometric stability.

1.4 Publications and Patents

The work presented in this thesis has been published in the following journal and conference papers.

ECG recognition algorithms and frameworks (enhanced AC/LDA)

- F. Agrafioti, D. Hatzinakos, Signal Validation for Cardiac Biometrics, *IEEE 35th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010)*, March 14-19, 2010, Dallas, Texas, USA

- F. Agrafioti, J. Gao, D. Hatzinakos, Heart Biometrics: Theory, Methods and Applications, in Biometrics: Book 3, J. Yang, Eds., Intech (In publication)

- J. Gao, F. Agrafioti, H. Mohammadzade, D. Hatzinakos, ECG for Blind verification in Distributed Systems, *IEEE 36th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, May 22-27, Prague

- F. Bui, F. Agrafioti, and D. Hatzinakos, Electrocardiogram (ECG) biometric for robust identification and secure communication , in *Biometrics: Theory, Methods and Applications*, N. Boulgouris, E. Micheli-Tzanakou, and K. Plataniotis, Eds. Wiley

- F. Agrafioti, F. M. Bui, D. Hatzinakos, Medical Information Management with ECG Biometrics: A Secure and Effective Framework, in *Handbook on Ambient Assisted Living for Healthcare, Well-being and Rehabilitation*, Paul McCullagh, IOS Press (In publication)

Affective Computing using the ECG

- F. Agrafioti, D. Hatzinakos, A. K. Anderson, ECG Pattern Analysis for Emotion Detection, *IEEE Transactions on Affective Computing*, 14 pages, submitted January 2011 (Under second review with minor revisions)

- F. Agrafioti, D. Hatzinakos, An Enhanced EMD Algorithm for ECG Signal Processing, *IEEE 17th International Conference on Digital Signal Processing (DSP 2011)*, July 6-8, Corfu, Greece

ECG biometrics in monitoring

- F. Agrafioti, F. M. Bui, D. Hatzinakos, Medical Biometrics in Mobile Health Monitoring, *Wiley's Security and Communication Networks Journal, Special Issue on Biometric Security for Mobile Computing*, vol. 4, no. 5, pp. 525-539, July 2010

- F. Agrafioti, F. M. Bui, D. Hatzinakos, Medical Biometrics: The Perils of Ignoring Time Dependency, *IEEE Third International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*, Sept. 28-30, 2009, Washington DC, USA

ECG biometrics in anonymous frameworks

- F. Agrafioti, F. M. Bui, D. Hatzinakos, On Supporting Anonymity in a BAN Biometric Framework, *IEEE 16th International Conference on Digital Signal Processing (DSP 2009)*, July 5-7, 2009, Santorini, Greece

Patent Filing of the following patent is anticipated:

- F. Agrafioti, F. M. Bui, and D. Hatzinakos, "ECG-Based Recognition Frameworks for Small and Large Scale Applications", US Provisional Patent # 61484470

Note that the proposed frameworks rely heavily on this author's prior work on ECG biometrics as well as on related publications on the Empirical Mode Decomposition, published in:

- F. Agrafioti, J. Gao, H. Mohammadzade, D. Hatzinakos, A 2D Bivariate EMD Algorithm for Image Fusion, *IEEE 17th International Conference on Digital Signal Processing (DSP 2011)*, July 6-8, Corfu, Greece

- H. Mohammadzade, F. Agrafioti, J. Gao, D. Hatzinakos, BEMD for Expression Transformation in Face Recognition, *IEEE 36th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, May 22-27, Prague

- Z. S. Fatemian, F. Agrafioti, D. Hatzinakos, HeartID: Cardiac Biometric Recognition , *IEEE Fourth International Conference on Biometrics: Theory, Applications and Systems (BTAS 2010)*, Sept. 27-29, 2010, Washington DC, USA

- F. Agrafioti, D. Hatzinakos, ECG Biometric Analysis in Cardiac Irregularity Conditions , *Signal, Image and Video Processing, Springer*, vol. 3, no. 4 pp 329-343, 2009

- F. Agrafioti, D. Hatzinakos, Fusion of ECG sources for human identification, *IEEE 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP 2008)*, March 12-14, 2008, Malta

- F. Agrafioti, D. Hatzinakos, ECG based recognition using second order statistics, *IEEE 6th Annual Conference on Communication Networks and Services Research (CNSR 2008)*, May 5-8, 2008, Halifax, Canada

- Y. Wang, F. Agrafioti, D. Hatzinakos and K. N. Plataniotis, Analysis of Human Electrocardiogram for Biometric Recognition, *EURASIP, Journal on Advances in Signal Processing, Special Issue on Advanced Signal Processing and Pattern recognition Methods for biometrics*, Article ID 148658, May 2007.

1.5 Thesis Outline

The remainder of this thesis is organized as follows. Chapter 2 covers the background information and prior art in the areas of ECG biometric recognition as well as affective computing. Appendix B contains a detailed discussion on emotion modeling. Chapter 3 presents the experimental protocols that have been designed for ECG data collection. Chapter 4 describes the proposed algorithm for ECG quality assessment and feature extraction (*HeartID*). This chapter also presents the application frameworks for this technology along with the customization of the recognition algorithm for every setting. Simulation results on biometric recognition are reported at the end of this chapter.

Chapter 5 presents the analysis for emotion detection from ECG signals. This analysis relies heavily on the Empirical Mode Decomposition, a description of which can be found in Appendix A. In addition, chapter 5 presents the performance of the emotion classification system evaluated over ECG signals acquired with customized experimental setups. Chapter 6 brings the reader back to the biometric recognition topic by incorporating the ECG affective computing findings into the *HeartID* system. The thesis concludes with Chapter 7.

Chapter 2

Background and Prior Art

2.1 Taxonomy of Errors in Biometric Recognition

Matching two biometric feature vectors does not have a single positive or negative answer. In biometrics, even though features originate from the same subject, significant intra-class variability is usually observed that renders classification very difficult. The error of such systems is directly linked to the mode under which they operate. For this reason, the biometric modes of operation are herein presented first.

1. **Enrollment.** During this mode of operation the biometric system collects the recognizing modality (ex. ECG, face, iris), performs some quality check, processes to extract discriminative features, and stores the result in the *gallery set*.

2. **Identification.** During this mode of operation, the system uses an input biometric reading to perform *one-to-many* matches with the gallery set. The purpose of this operation is to answer the question: *What is the identity of this user?*

3. **Verification (or Authentication).** During this mode of operation, the user submits to the system a biometric sample along with an identity claim. The system compares the input sample with the corresponding record from the gallery set, and

a Good/Poor match decision is returned. The purpose of this operation is to answer the question: *Is the user who he/she claims to be?*

Instead of the credential information, the output of a biometric system is often a match score, revealing the degree of resemblance for a given pair of biometric samples. In essence, the match score expresses the degree of certainty (or uncertainty) about a user's identity. Match scores can take the form of a probability, similarity, or distance, and authentication is then carried out by setting a threshold empirically.

In order to distinguish the types of errors that a biometric system can make, it is important to outline the following states as the system's possible conditions:

1. *Identify an individual correctly*, which is measured in identification rates.
2. *Misidentify an enrolled individual*, which is measured in mis-identification rates.
3. For more complex systems, *authentication of legitimate subjects* is referred to as *sensitivity* and measured in authentication (or verification) rates.
4. *Deny identity authentication to a legitimate subject*, measured in false rejection rates (FRR).
5. *Deny identity authentication to intruders*, referred to as *specificity* of the system.
6. *Authenticate intruders*, which is measured in false acceptance rates (FAR).

Specifically, the false acceptance and rejection statistics are computed as:

$$\text{FAR} = \frac{\text{Number of falsely authenticated subjects}}{\text{Total number of intruders}} \tag{2.1}$$

$$\text{FRR} = \frac{\text{Number of rejected legitimate subjects}}{\text{Total number of ID attempts}} \tag{2.2}$$

The equal error rate (EER) is also defined as the point in the FAR and FRR curves, where false acceptance is equal to false rejection i.e., $EER = FAR = FRR$. The lower the equal error rate, the better the authentication performance of the system.

Depending on the employed similarity measure, the appearance of the FAR and FRR distribution can vary. When distance is used to associate two records, suggesting that the higher the score the less the resemblance, FAR is expected to increase as the distance threshold increases. This way, for a higher selection of the threshold, intruder authentication is rendered more likely. Similarly, the false rejection percentage is expected to fall as the distance threshold increases, because more legitimate subjects will be rejected. Obviously, there is a trade-off between false acceptance and rejection cases, and even though ideally a biometric system would demand both of them to be low, it is usually left up to the designer to decide on the specifics of the application.

2.2 Electrocardiogram Fundamentals

The electrocardiogram (ECG) is one of the most widely used signals in healthcare. Recorded at the surface of the body, with electrodes attached in various configurations, the ECG signal is studied for diagnostics even at the very early stage of a disease. In essence, this signal describes the electrical activity of the heart over time, and pictures the sequential depolarization and repolarization of the different muscles that form the myocardium.

The first recording device was developed by the physiologist Williem Einthoven in the early 20th century, and for this discovery he was rewarded with the Nobel Prize in Medicine [10]. Since then ECG became an indispensable tool in clinical cardiology. However, the deployment of this signal in biometric recognition and affective computing is relatively young.

Figure 2.1 shows the salient components of an ECG signal i.e., the P wave, the QRS

complex and the T wave. The P wave describes the depolarization of the right and left atria. The amplitude of this wave is relatively small, because the atrial muscle mass is limited. The absence of a P wave typically indicates ventricular ectopic focus. This wave usually has a positive polarity, with a duration of approximately 120 ms, while its spectral content is limited to 8-10 Hz, i.e., low frequencies.

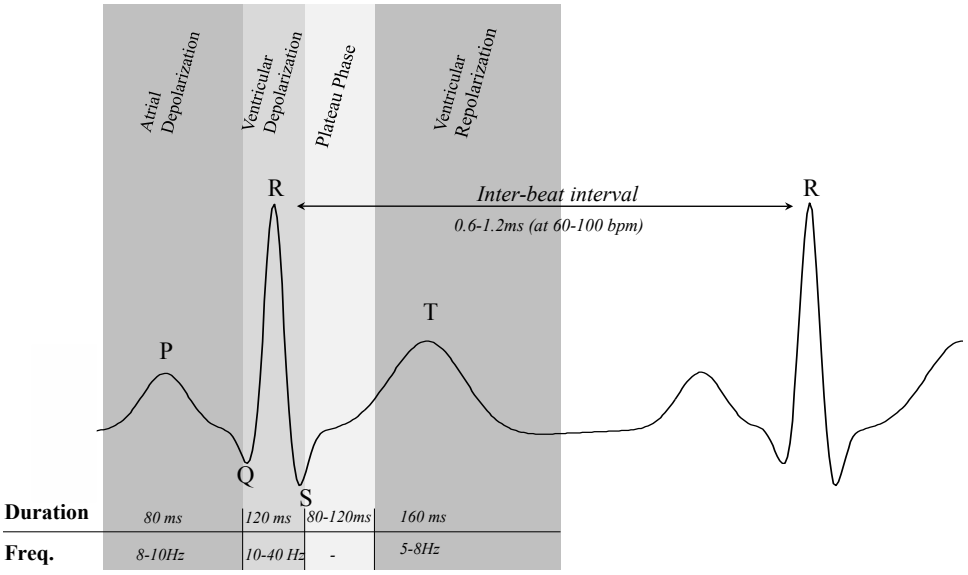


Figure 2.1: Main components of an ECG heart beat. Each wave describes a distinct phase of the cardiac cycle.

The QRS complex corresponds to the largest wave, since it represents the depolarization of the right and left ventricles, which are the chambers with the most substantial mass in the heart. The duration of this complex is approximately 70-110 ms in a normal heartbeat. The anatomic characteristics of the QRS complex depend on the origin of the pulse. Due to its steep slopes, the spectrum of a QRS wave is higher compared to that of other ECG waves, and is mostly concentrated between 10 and 40 Hz.

Finally, the T wave depicts the ventricular repolarization. It has a relatively small amplitude and is usually observed about 300 ms after the QRS complex. However, its precise position depends on the heart rate, e.g., appearing closer to the QRS wave at rapid heart rates.

There is more than one approach to ECG recording, such as the orthogonal leads and synthesized leads [10]. However, the most widely applied system is the *standard 12-lead ECG* where there are three main sets of lead orientations. The *bipolar limb leads* are usually denoted as I, II and III and they track the electrical potential of the heart when three electrodes are attached at the right and left wrist and left ankle [10].

By convention, lead I measures the potential difference between the two arms. In lead II, one electrode is attached on the left leg and the other one on the right hand as depicted in Figure 2.2. Finally, in lead III configuration, the measured potential is between the left leg and right hand.

Following the electrode position as pictured in Figure 2.2, the limb leads are measured in the following combinations:

$$I = V_{LH} - V_{RH} \tag{2.3}$$

$$II = V_{LL} - V_{RH} \tag{2.4}$$

$$III = V_{LL} - V_{LH} \tag{2.5}$$

The preceding equations suggest that, having recorded any two of the bipolar limb lead signals, the third one can be directly derived. The *augmented unipolar limb leads* fill the 60^o gaps in the directions of the bipolar limb leads. Using the same electrodes, the augmented unipolar leads are measured as:

$$aVR = V_{RH} - \frac{V_{LH} + V_{LL}}{2} \tag{2.6}$$

$$aVL = V_{LH} - \frac{V_{RH} + V_{LL}}{2} \tag{2.7}$$

$$aVF = V_{LL} - \frac{V_{LH} + V_{RH}}{2} \tag{2.8}$$

The third category of lead orientation involved in the conventional 12-lead system comprises the *precordial leads* (V1, V2, V3, V4, V5, V6). These signals are recorded with 6 electrodes attached successively on the left side of the chest, thus capturing more detailed information in the electrocardiogram [10].

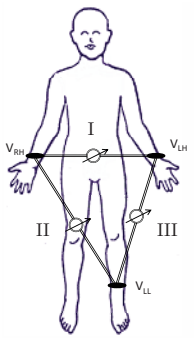


Figure 2.2: Configuration of Leads I II and III.

Apart from the well known physiological process that generates the ECG, the signal is also affected by various sympathetic and parasympathetic processes. From a signal processing point of view, this is the reason why the ECG signal is not perfectly periodic. In the subsequent discussion the purpose is to differentiate between the anatomical properties of the heart, which encourage its deployment in biometrics, and the sources of variability due to the autonomic nervous system that support its analysis in affective computing.

2.2.1 Inter-individual variability

This section will briefly discuss the physiological rationale for the use of ECG in biometric recognition. Overall, healthy ECG signals from different people conform to roughly the same repetitive pulse pattern. However, further investigation of a person's ECG signal can reveal notably unique trends which are not present in recordings from other individuals. The inter-individual variability of ECG has been extensively reported in the literature [11, 12, 13, 14, 15, 16, 17].

More specific, the ECG signal depicts the various electrophysiological properties of the cardiac muscle. Model studies have shown that physiological factors such as the heart mass orientation, the conductivity of various areas, and the activation order of the heart, are sources of significant variability among individuals [16, 17].

Furthermore, geometrical attributes such as the exact position and orientation of the myocardium, and torso shape designate ECG signals with particularly distinct and personalized characteristics. Other factors affecting the ECG signal are the timing of depolarization and repolarization and lead placement. In addition, except for the anatomic idiosyncrasy of the heart, unique patterns have been associated to physical characteristics such as the body habitus and gender [11, 15, 16, 17, 18]. The electrical map of the area surrounding the heart may also be affected by variations of other organs in the thorax [17].

In fact, various methodologies have been proposed to eliminate the differences among ECG recordings. The idea of clearing off the inter-individual variability is typical when seeking to establish healthy diagnostic standards [12]. Automatic diagnosis of pathologies using the ECG is infeasible if the level of variability among healthy people is high [16]. In such algorithms, personalized parameters of every subject are treated as random variables and a number of criteria have been defined to quantify the degree of subjects' similarities on a specific feature basis.

2.2.2 Cardiovascular reactivity to emotion

The central and peripheral nervous systems (CNS, PNS) of the human body are responsible for physical and psychological behavior. The CNS is located in the cranial cavity and the spinal cord, and is the information processing unit of the body. The PNS acts as the communication channel between the CNS and the organs, and consists of information transferring nerves. Depending on whether the individual has control over these nerves or not, the PNS is divided into the *somatic* and the *autonomic nervous systems* (SNS, ANS). Various essential involuntary activities of the human body such as heart pulses, respiration, salivation etc. come under the ambit of the ANS. The ANS is further classified into the *sympathetic* and *parasympathetic* subsystems, which operate in an antagonistic manner. The resulting manifestation depends on which subsystem is dominant

at a particular instance. The opposing nature of the two results in a mannered balance within the body, called *homeostasis*.

The ANS nerve-endings within the cardiac muscle play a major role in the cardiac output because they affect the rhythm at which the muscle pumps blood. The fibers of the sympathetic system run along the atria and the ventricles, and when activated stimulate the cardiac muscle to increase the heart rate. On the other hand, the parasympathetic system reduces the cardiac workload. Specifically, in the presence of a mental stressor, the sympathetic system dominates the parasympathetic, resulting in the following reactivity effects [19]:

1. *Automaticity.* The intrinsic impulse firing (automaticity) of the pacemaker cells increases, which translates directly to an increased heart rate.

2. *Contractility.* During every contraction the fibers of the heart shorten more, compared to the case during homeostasis, thereby increasing the force of contraction.

3. *Conduction rate.* The natural pacemaker, the SA node, is forced to conduct faster.

4. *Excitability.* During sympathetic stimulation, the person has increased perceptiveness to internal and external stimuli, which increases the irritability of the cardiac muscle and possibly lead to ectopic beats.

5. *Dilation of coronary blood vessels.* The diameter of the coronary blood vessels increases, followed by increased blood flow to the cardiac muscle.

Depending on the intensity of a particular emotion, the sympathetic system is stimulated to prepare the body for *vigorous* activity. Apart from the well established cardiac reaction to ANS, there, there are questions that are yet to be answered. For instance, it is not clear whether the sympathetic and parasympathetic branches of ANS affect the heart in the same way i.e., whether there exist independent ANS drives on the cardiac muscle or not.

Shouldice *et al.* [20] provided evidence that there is separate ANS innervation to the two pacemakers (SA and VA nodes). By analyzing PP and PR intervals in transitions from supine to stand postures and vice versa, the authors showed a decoupling of the SA and VA modulation. This suggests that the cardiac reactivity may contain more detail about the stimulus than just the intensity (arousal). Another interesting finding in [20] is that there is large inter-subject autonomic innervation variability, meaning that one cannot generalize how the excitability of the SA and VA nodes is affected over a population. Their work leads one to suspect the specificity of emotion with regard to particular people.

Furthermore, signals such as the heart rate variability (HRV) and blood volume pressure (BVP) have been associated with the development of coronary artery disease. The cardiac reactivity to stress has been studied in relation to feelings such as anxiety, hostility or challenge [21, 22]. Despite the dependence of the HRV time series on posture (standing/supine), increased vagal modulation of the cardiovascular system was statistically correlated with the aforementioned emotions. Similarly, the difficulty levels of memory tasks were shown to impact cardiac reactivity [23]. Furthermore, Blascovich *et al.* [24] were able to differentiate challenge from threat using cardiac traits such as the HR, ventricular contractility and cardiac output.

As already discussed, the most widely studied trait of cardiovascular activity is the HR, or HRV [25, 26, 27]. Based on the effects of sympathetic stimulation on the cardiac muscle, the HR is the most natural choice for arousal detection using comparison of sympathetic and parasympathetic frequency bands of the time series [27]. However, it is highly dependent on the position of the body during monitoring [28, 21].

Emotion specific reactivity has been observed on the ECG signal itself, without collapsing the embedded information to arbitrary interval measurements (HRV). In fact, Andrassy *et al.* [29] observed QT prolongation (typical for stress) without significant R-R interval changes.

Folino *et al.* [30] recorded measurements of features such as the duration of the QRS complex, low amplitude waves at the terminal portions of the complex, and the root mean square voltage of the QRS while the subject was performing various mental arithmetic tasks. They observed that with increasing levels of difficulty, the energy of the QRS complex increased significantly, while its duration was reduced (not a result of increasing HR).

The T wave amplitude was cross examined with mental stress in [31, 32]. An interesting observation brought to light by these works *was that manifestations of mental stress on ECG depends on whether the stressor is active or passive.* Contrary to the popular belief that mental stress response is generic, no significant effects were observed during passive tasks (i.e. stressors without the active involvement of the subject).

Apart from mental stress, cardiovascular responses have been examined for other particular emotions. Sinha *et al.*[33] experimented with five emotional states namely joy, sadness, fear, anger and neutral. The HR, BVP, stroke volume, pre-ejection period and cardiac output were measured, employing imagery tasks for emotion elicitation. Pairwise comparison performance (for every two emotions) was reported for every measurement, with the HR being among the signals that exhibit the least discriminative power.

In summary, there is significant evidence in the literature that cardiovascular reactivity can differentiate not only the intensity of the stimulus but also the valence (see Appendix B), depending on how the stimulus is presented. Although in the affect research ECG is primarily used for HR estimation, we argue that it is not used to its full extend, as this signal is yet one of the most illustrative and detailed records of cardiac activity. Furthermore, ECG exhibits idiosyncratic patterns due to the unique anatomy of people's cardiac muscles. Along these lines, in this work we investigate the dynamics of the ECG signal and their association with emotion specific ANS activity, from a subject specific point of view.

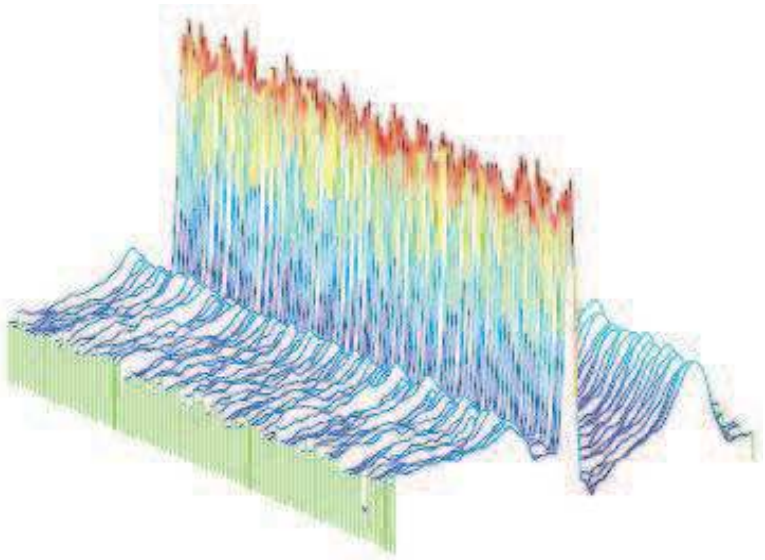


Figure 2.3: Variability surrounding the *QRS* complex among heart beats of the same individual.

2.3 ECG Biometric Recognition: Literature Survey

Prior works in the ECG biometric recognition field can be categorized as either fiducial points dependent or independent. Fiducials are specific points of interest on an ECG heart beat such as the ones shown in Figure 2.1. Fiducial based approaches rely on local features of the heart beats for biometric template design, such as the temporal or amplitude difference between consecutive fiducial points. On the other hand, fiducial points independent approaches treat the ECG signal or isolated heart beats holistically and extract features statistically based on the overall morphology of the waveform. This distinction has a direct analogy to face biometric systems, where one can operate locally and extract biometric features such the distance between the eyes or the size of the mouth. Alternatively, a holistic approach would be to analyze the facial image globally.

Both approaches have advantages and disadvantages. While fiducial oriented features risk to miss identifying information hidden behind the overall morphology of the biometric modality, holistic approaches deal with a large amount of redundant information that

needs to be eliminated. The challenge in the later case, is to remove this information in a way that the intra-subject variability is minimized, and the inter-subject is maximized. For the ECG case, detecting fiducial points is a very obscure process due to the high variability of the signal. Figure 2.3 shows an example of aligned ECG heart beats which belong to the same individual. Even though the QRS complex is perfectly aligned, there is significant variability surrounding the P and the T waves which renders the localization of these waves' onsets and offsets very difficult. In fact, there is no universally acknowledged rule that can guide this detection [17].

This section provides an overview of fiducial dependent and independent approaches that are currently found in the literature. A comparison is also provided in Tables 2.1 and 2.2.

2.3.1 Fiducial Based Approaches

Among the earliest works in the area is Biel *et al.*'s [34] proposal, in 2001, for a fiducial feature extraction algorithm, which demonstrated the feasibility of using ECG signals for human identification. The standard 12 lead system was used to record signals from 20 subjects of various ages. Special experimentation was carried out to test variations due to lead placement in terms of the exact location and the operators who carry out the procedure.

Out of 30 clinical diagnostic features that were estimated for each of the 12 leads, only 12 features were retained for matching, by inspection of the correlation matrix. These features pictured local characteristics of the pulses, such as the QRS complex and T wave amplitudes, P wave duration and other. This feature set was subsequently fed to SIMCA for training and classification. Results of combining different features were compared to demonstrate that, in the best case, classification rate was 100% with just 10 features.

Kyoso *et al.*, [35], also proposed fiducial based features for ECG biometric recognition.

Overall, four feature parameters were selected i.e., the P wave duration, PQ interval, QRS complex and QT durations. These features were identified on the pulses by applying a threshold to the second order derivative. The subject with the smallest Mahalanobis distance between each two of the four feature parameters was selected as the output. The highest reported performance was 94.2% for using just the QRS and QT intervals.

In 2002, Shen *et al.* [36] reported an ECG based recognition method with seven fiducial based features that relate to the QRS complex. The underlying idea was that this wave is less affected by varying heart rates, and thus is appropriate for biometric recognition.

The proposed methodology encompassed two steps [36]: During a first step, template matching was used to compute the correlation coefficient among the QRS complexes in the gallery set in order to find possible candidates and prune the search space. A decision based neural network (DBNN) was then formed to strengthen the validation of the resulting identity. While the first step managed to correctly identify only 85% of the cases, the neural network resulted in 100% recognition.

More complete biometric recognition tests were reported in 2004, by Israel *et al.* [37]. This work presented the three clear stages of ECG biometric recognition i.e., preprocessing, feature extraction and classification. In addition, a variety of experimental settings are described in [37] such as, examination of variations due to electrode placement and physical stress.

The proposed system employed only temporal features. A filter was first applied to retain signal information in the band 1.1- 40 Hz and discard the rest of the spectral components which were attributed to noise. By targeting to keep discriminative information while applying a stable filter over the gallery set, different filtering techniques were examined to conclude to a local averaging, spectral differencing and Fourier band-pass filter. The highest identification rate achieved was close to 100% which generally established the ECG signal as a biometric modality.

A similar approach was reported in the same year by Palaniappan *et al.*, [38]. In addition to commonly used features within the QRS complex, a form factor, which is a measure of signal complexity, was proposed and tested as input to a neural network classifier. An identification rate of 97.6% was achieved over recordings of 10 individuals, by training a MLP-BP with 30 hidden units.

Kim *et al.* [39], proposed a method to normalize time domain features by up-sampling the heart beats. In addition, the P wave was excluded when calculating the features, since it disappears when heart rate increases. With this strategy, the performance improved significantly when the testing subjects were performing physical activities.

Another work that addressed the heart rate variations was by Saechia *et al.*, [40] in 2005. The heart beats were normalized to healthy durations and then divided into three sub-sequences: P wave, QRS complex and T wave. The Fourier transform was applied on a heart beat itself and all three sub-sequences. The spectrums were then passed to a neural network for classification. It was shown that false rate was significantly lower (17.14% to 2.85%) by using the three sub-sequences instead of the original heart beat.

Zhang *et al.* [41], suggested 14 commonly used features from ECG heart beats on which a PCA was applied to reduce dimensionality. A classification method based on the Bayes Theorem was proposed to maximize the posterior probability given prior probabilities and class-conditional densities. The proposed method outperformed Mahalanobis' distance by 3.5% to 13% depending on the particular lead that was used.

Singh *et al.* [42], proposed a way to delineate the P and T waveforms for accurate feature extraction. By examining the ECG signal within a preset window before the Q wave onset and apply a threshold to its first derivative, the precise position of the P was revealed. In addition to the onset, the peak and offset of the P wave were detected by tracing the signal and examining the zero crossings of its first derivative. The accuracy of this system was reported as 99%, tested over 25 subjects.

In 2009, Boumbarov *et al.* [43], investigated different models such as HMM-GMM

(Hidden Markov model with Gaussian mixture model), HMM-SGM (Hidden Markov model with single Gaussian model) and CRF (Conditional Random Field), to determine different fiducial points in an ECG segment, followed by PCA and LDA for dimensionality reduction. A neural network with a radial basis function was realized as the classifier and the recognition rate was between 62% to 94% for different subjects.

Ting *et al.* [44], described in 2010 a nonlinear dynamical model to represent the ECG in a state space form with the posterior states inferred by an Extended Kalman Filter. The Log-likelihood score was used to compare the estimated model of a testing ECG to that of the enrolled templates. The reported identification rate was 87.5% on the healthy beats of 13 subjects from the MIT Arrhythmia database. It was also reported that the method was robust to noise for SNR above 20 dB.

The Dynamic time warping or FLDA were used in [45], together with the nearest neighbor classifier. The proposed system was comprised of two steps as follows: First the FLDA and nearest neighbor operated on the features and then a DTW classifier was applied to additionally boost the performance (100% over a 12-subject database). For verification, only features related to QRS complex were selected due to their robustness to heart rate variability. The same two-stage setting was applied together with a threshold and the reported performance was 96% for 12 legitimate subjects and 3 intruders.

Another fiducial based method was proposed by Tawfik *et al.* [46]. In this work, the ECG segment between the QRS complex and the T wave was first extracted and normalized in the time domain by using Framingham correction formula or by assuming constant QT interval. The DCT was then applied and the coefficients were fed into a neural network for classification. The identification rate was 97.72% for that of the Framingham correction formula and 98.18% for the case of constant QT interval. Interestingly, using only the QRS complex without any time domain manipulation yielded a performance of 99.09%.

In summary, although a number of fiducial based approaches have been reported for

ECG based biometrics, accurate localization of fiducial points remains a big challenge. This ambiguity risks the accuracy of the respective recognizers which require the precise location of such points. In the likely event of failing to adequately determine the locations of these points, fiducial approaches would rather reject the heart beat and require an extra reading, rather than risking the accuracy of their decision. This however, results in increased rejection rates that subsequently increase thw waiting time of the system. It is important to note, that in order to to acquire one clear pulse the sensors need to be attached/ held by the user for some time, until no muscular movements take place. Therefore, additional recording sessions due to rejection increase the inconvenience of use of such systems.

2.3.2 Fiducial Independent Approaches

On the non-fiducial methodologies side, the majority of the works were reported after 2006. Among the earliest is Plataniotis *et al.*'s [47] proposal for an autocorrelation (AC) based feature extractor. With the objective of capturing the repetitive pattern of ECG, the authors suggested the AC of an ECG segment as a way to avoid fiducial points detection. It was demonstrated that the autocorrelation of windowed ECG signals embeds highly discriminative information in a population. However, depending on the original sampling frequency of the signal, the dimensionality of a segment from the autocorrelation was considerably high for cost efficient applications. To reduce the dimensionality the discrete cosine transform (DCT) was applied. The method was tested on 14 subjects, for which multiple ECG recordings were available, acquired a few years apart. The identification performance was 100%. However, the DCT reduces the dimensionality from a signal processing point of view, which is sub-optimal for classification problems.

Wubbeler *et al.* [48], have also reported an ECG based human recognizer by extracting biometric features from a combination of Leads I, II and III i.e., a two dimensional heart vector also known as the characteristic of the ECG. To locate and extract pulses a thresh-

olding procedure was applied. For classification, the distance between two heart vectors as well as their first and second temporal derivatives were calculated. A verification functionality was also designed by setting a threshold on the distances. Authenticated pairs were considered those which were adequately related, while in any other case, input signals were rejected. The reported false acceptance and rejection rates were 0.2% and 2.5% corresponding to a 2.8% equal error rate (EER). The overall recognition rate of the system was 99% for 74 subjects.

A methodology for ECG synthesis was proposed in 2007 by Molina *et al.* [49]. A heart beat was normalized and compared with its estimate, which was previously constructed from itself and the templates from a claimed identity. The estimated version was produced by a morphological synthesis algorithm involving a modified dynamic time warping procedure. The Euclidean distance was used as the similarity measure and a threshold was applied to decide the authenticity. The highest reported performance was 98% with a 2% EER.

In 2008, Chan *et al.* [50], reported ECG signal collection from the fingers by asking the participants to hold two electrode pads with their thumb and index finger. The Wavelet distance was used as the similarity measure with a classification accuracy of 89.1%, which outperformed other methods such as the percent residual distance and the correlation coefficient. Furthermore, an additional recording session was conducted for several misclassified subjects which overall improved the system's performance to 95%.

In the same year, Chiu *et al.* [51], proposed the use of DWT on heuristically isolated pulses. More precisely, every heart beat was determined on the ECG signal, as 43 samples backward and 84 samples forward from the R peaks. The DWT was used for feature extraction and the Euclidean distance as the similarity measure. When the proposed method was applied to a database of 35 healthy subjects, a 100% verification rate was reported. The author also pointed out that false rate would increase if 10 subjects with cardiac arrhythmia were included in the database.

Fatemian *et al.* [52], also suggested the Wavelet transform to denoise and delineate the ECG signals, followed by a process wherein every heart beat was resampled, normalized, aligned and averaged to create one strong template per subject. A correlation analysis was directly applied to test heart beats and the template since the gallery size was greatly reduced. The reported recognition rate was 99.6% for a setting where every subject has 2 templates in the gallery.

The Spectogram was employed in [53] to transform the ECG into a set of time-frequency bins which were modeled by independent normal distributions. Dimensionality reduction was based on Kullback-Leibler divergence where a feature is selected only if the relative entropy between itself and the nominal model (which is the spectrogram of all subjects in database) is larger than a certain threshold. The log-likelihood ratio was used as a similarity measure for classification and different scenarios were examined. For enrollment and testings over the same day, a 0.37% ERR was achieved for verification and a 99% identification rate. For different days, the respective performance was 5.58% ERR and 76.9%.

Ye *et al.* [54], applied the discrete wavelet transform (DWT) and independent component analysis (ICA) on ECG heart beat segments to obtain 118 and 18 features respectively. The feature vectors were concatenated. The dimensionality of the feature space was subsequently reduced from 136 to 26 using PCA which retained 99% of the data's variance. An SVM with Guassian radial basis function was used for classification with a decision level fusion of the results from the two leads. A rank-1 classification rate of 99.6% was achieved for healthy heart beats. Another observation was that even though dynamic features such as the R-R interval proved to be beneficial for arrhythmia classification, they were not as good descriptors for biometric recognition.

Coutinho *et al.* [55], isolated the heart beats and performed an 8-bit uniform quantization to map the ECG samples to strings from a 256-symbol alphabet. Classification was based on finding the template in the gallery set that results in the shortest description length of the test input (given the strings in the particular template) which was calculated by the Ziv-Merhav cross parsing algorithm. The reported classification accuracy was 100% on a 19-subject database, in presence of emotional state variation.

Autoregressive modeling was used in [56]. The ECG signal was segmented with 50% overlap and an AR model of order 4 was estimated so that its coefficients are used for classification. Furthermore, the mean PSD of each segment was concatenated as add-on features which increased the overall performance to 100% using a k-Nearest Neighbor classifier.

Li *et al.* [57], proposed a method to model the ECG signal in both the temporal and cepstral domain. The hermite polynomial expansion was employed to transform heart beats into Hermite polynomial coefficients which were then modeled by an SVM with a linear kernel. Cepstral features were extracted by simple linear filtering and modeled by GMM/GSV (GMM super-vector). The highest reported performance was 98.26% with a 5% ERR corresponding to a score level fusion of both temporal and cepstral classifiers.

It is clear from the above, that a large variety of fiducial independent techniques have been proposed for ECG biometric analysis. While some approaches are more computational intensive than other, or they operate on heart beats rather than finite ECG segments, there are practically a number of open issues in the literature with regard to ECG biometrics.

Among the most prominent ones is the question of signal stability, or permanence, with time. The majority of prior works did not examine the evolution of the ECG signal with time. To some extent, the sources of intra-subject variability of the ECG signal have been ignored. We advocate that the study of the factors that affect the ECG waveform and may render the biometric template less accurate is very important for real

Method	Principle	Performance	Number of subjects
Kyoso *et. al* [35]	Analyzed four fiducial based features from heart beats, to determine those with greater impact on the identification performance	94.2%	9
Biel *et. al* [34]	Use a SIEMENS ECG apparatus to record and select appropriate medical diagnostic features for classification	100%	20
Shen *et. al* [36]	Use template matching and neural networks to classify QRS complex related characteristics	100%	20
Israel *et. al* [37]	Analyze fiducial based temporal features under various stress conditions	100%	29
Palaniappan *et. al* [38]	Use two different neural network architectures for classification of six QRS wave related features	97.%	10
Kim *et. al* [39]	By normalizing ECG heartbeat using Fourier synthesis the performance under physical activities was improved	N/A	10
Saechia *et. al* [40]	Examined the effectiveness of segmenting ECG heartbeat into three subsequences	97.15%	20
Zhang *et. al* [41]	Bayes' classifier based on conditional probability was used for identification and was found supurior to Mahalanobis' distance.	97.4%	502
Plataniotis *et. al* [47]	Analyze the autocorrelation of ECGs for feature extraction and apply DCT for dimensionality reduction	100%	14
Wubbeler *et. al* [48]	Utilize the characteristic vector of the electrocardiogram for fiducial based feature extraction out of the QRS complex	99%	74

Table 2.1: Summary of related to ECG based recognition works

Molina *et. al* [49]	Morphological synthesis technique was proposed to produced a synthesized ECG heartbeat between the test sample and template	98%	10
Chan *et. al* [50]	Wavelet distance measure was introduced to test the similarity between ECGs	95%	50
Singh *et. al* [42]	A new method to delineate P and T waves	99%	25
Fatemian *et. al* [52]	Less templates per suject in gallery set to speed up computation and reduce memory requirement	99.6%	13
Boumbarov *et. al* [43]	Neural network with radial basis function was employed as the classifier	86.1%	9
Ting *et. al* [44]	Use extended Kalman filter as inference engine to estimate ECG in state space	87.5%	13
Odinaka *et. al* [53]	Time frequency analysis and relative entropy to classify ECGs	76.9%	269
Venkatesh *et. al* [45]	Apply dynamic time warping and Fisher's discriminant analysis on ECG features	100%	15
Tawfik *et. al* [46]	Examined the system performance when using normalized QT and QRS and using raw QRS	99.09%	22
Ye *et. al* [54]	Applied Wavelet transform and Independent component analysis, together with support vector machi as classifier to fuse information from two leads	99.6%	36 normal and 112 arrhythmic
Coutinho *et. al* [55]	Treat heartbeats as a strings and using Ziv-Merhav parsing to measure the cross complexity	100%	19
Ghofrani *et. al* [56]	Autoregressive coefficient and mean of power spectral density were proposed to model the system for classification model the system for classification	100%	12
Li *et. al* [57]	Fusion of temporal and cepstral features	98.3%	18

Table 2.2: Summary of related to ECG based recognition works (Continued)

life deployment of this technology.

Furthermore, the one-dimensional representation of the ECG signal presents significant advantages to fast identity computation, and complex processing algorithms (neural networks, warping etc.) may compromise this asset. Overall, the prior art in ECG biometrics recognition, as summarized in Tables 2.1 and 2.2, did not discuss the way that the respective solutions will fit in real-life settings.

2.4 ECG in Affective Computing: Literature Survey

Emotion recognition in the field of human computer interaction has drawn enormous attention as early as the 1980s. To date, there are numerous works in literature focusing on facial expression or speech analysis for the purpose of emotion detection. Irrespectively of the employed modality, each reported work contributes to the overall understanding of emotions and to the way they are expressed through the human body. Nevertheless, this section places the emphasis on physiological signal processing for detection and classification of emotions.

Although the approaches found in literature vary algorithmically and with respect to the experimental procedures employed, current methodologies can be classified in to two categories based on the way the emotional models are conceived and deployed. The first category includes *discrete emotional models* (DEM), where the main objective of the proposed frameworks is to recognize and label standard emotional states (for instance joy, sadness or fear) depending on the application. These methods rely heavily on the assumption that any affective state is physiologically well defined and completely distinguishable from the rest.

On the other hand, the *affective dimensional models* (ADM) relax the conditions for discrete emotions and treat any affective state as a combination of two parameters, namely *arousal* and *valence*. Arousal is a measure of emotional stimulation which varying

from low to high. Valence is a measure of pleasantness for the experienced emotion, ranging from very pleasant to unpleasant. These two measures define a two dimensional space (the AV plane), and classification is carried out among predefined areas of the plane, which do not correspond necessarily to a discrete emotion. More information pertaining to emotion definitions can be found in Appendix B.

Both methodologies have advantages and disadvantages. For instance, the discrete emotional models might be easily conceptualized by people and fitted to applications, however their performance is risked as there is no well defined physiological borders among emotions. The rest of this section provides a literature survey for both discrete and dimensional emotional models.

2.4.1 Discrete Emotional Models

To the best of our knowledge, the first documented effort for emotion recognition using physiological signals is by Ekman *et al.* [58] in 1983. This work was considered a breakthrough as it opposed to the dogma dictates that no activity of the autonomic nervous system (ANS) is emotion specific. In other words, Ekman *et al.*'s work provided evidence that emotions can be distinguished among the various reactions of the human body.

Using both physiological signals and videos of facial expressions, six emotions were studied: surprise, disgust, sadness, anger, fear and happiness. The bio signals considered were the heart rate, left and right hand temperature, skin resistance and forearm muscle tension. The respective feelings were induced in a different manner for two recording modalities. For facial gestures, the subjects were asked to mimic emotional expressions that they had practiced before using a mirror, and following the instructions of an operator concerning the facial muscles they were required to contract.

On the other hand, for the stimulation of physiological signals an experience reliving approach was chosen. For this imagery task, every subject was asked to relax and bring to memory a personal experience that related to each of the studied emotions. Upon

completion, the subjects were asked to rank the intensity of their feelings on a predefined scale. For this experiment, 16 subjects were examined, 12 of which were professional actors. It is important to note that using actors to induce and analyze feelings, was a standard practice in the beginning of the affective computing research, as they were expected to manifest emotions naturally.

The signal analysis performed in [58], was based on the *change scores* principle. Data during emotional instances were averaged and subtracted from data of non-emotional phases. These differences were used as features which were then examined to evaluate organization in clusters based on tree decisions (for example, if the heart rate change was high and temperature was low, this would be assigned to the fear cluster). This technique although simplistic, demonstrated for the first time that physiological information can be organized in emotional clusters.

Ekman *et al.*'s findings have drawn the attention of the medical field (both physiological and psychological medicine), which has contributed significantly in the field. In 1992, Sinha *et al.* [33], reported a statistical analysis concerning the cardiovascular differentiation of emotions. Six heart related indexes were recorded from 27 male subjects. The signals were the heart rate, blood pressure, stroke volume, cardiac output, peripheral vascular resistance and an index of myocardial contractility. The objective was to examine and identify patterns of the cardiovascular activity under the expression of 4 main emotions: fear, anger, joy and sadness.

The experimental procedure in [33] involved a careful screening of the subjects participating in the study. This selection is generally prominent as the volunteers should be persons who not only experience various emotional states, but are also able to describe them verbally with accuracy. The physiological reaction of humans has been shown to be relative to their imagery ability [59]. For this reason, the subjects were selected with a questionnaire. The main experimental part involved two phases, one for the volunteers to familiarize with the imagery procedure (training), and one for the data collection.

The methodology for emotion analysis described in [33], resembles Ekman *et al.*'s proposal. Change scores were calculated for every cardiovascular measure by subtracting the baseline mean scores from the mean score during an emotional phase. With this kind of analysis it was demonstrated that the heart rate is very sensitive to anger compared the less valent feelings of joy or sadness. Another significant finding was that even though the emotions of fear and anger can increase the heart rate, systolic blood pressure and mascular blood flow, they can be differentiated by carefully examining the diastolic blood pressure, which decreases for fear only. In other words, peripheral vascular resistance decreases under fear. A similar study [60] in 1996 has expanded this analysis to non cardiovascular features, and analyzed skin conductance under mental stress to show very high correlation between this measure and a chosen psychological state.

In 2001 Picard (the pioneer of this field) *et al.* [1], proposed an emotionally intelligent system that utilizes information of 4 physiological signals. This work was the first to define and analyze the challenges of affective computing. The authors emphasize the fact that data collection for the purpose of emotion detection is a sensitive process compared to any standard experimental setup, and requires special attention. This is because in order to gather valid affective data, one has to make sure that first, the desired emotion was induced and second, that the subject labeled it accurately [61].

This is a non trivial procedure, first because as explained earlier some subjects are suppressive, thus any attempt to cause an emotional disturbance might fail, and second because people might have experienced feelings but not being aware of them or not feeling comfortable to declare them. Accordingly, Picard *et al.* [1] enumerated five parameters to be considered when designing such experiments:

1. An emotion can be self elicited (imagery), or controlled by external factors like audio and visual stimuli.

2. The experimental environment plays a major role and the level of familiarity with

Signal	Annotation	Sensor Placement
Electromyogram (EMG)	E	Masseter muscle (Face)
Blood Volume Pressure (BVP)	B	Tip of ring finger (left hand)
Skin Conductance	S	Index and middle fingers (left hand palm)
Respiration Rate	R	Diaphragm
Heart Rate	H	Measured from BVP peaks

Table 2.3: Biosignals analyzed in [1].

the subject should be defined.

3. There are several different modalities to explore (face, voice, gesture, biosignals).

4. The subject can be aware or unaware that some signal is being monitored.

5. The subject can be aware or unaware of the purpose of the experiment.

In addition, one of the major contributions in [1] is that data collection was carried out over multiple days for the same individual. This was novel not only from a data perspective, but also because it moved the interest of the community to examining also person-dependent recognition methodologies instead of modeling emotion by averaging population data. Furthermore, the stability of the emotional expression was examined over a period of time. There were 8 emotions under consideration: anger, hate, grief platonic love, romantic love, joy and reverence. Because this list has several non traditional emotions, the authors argued that the exact selection of emotions was not important, as the only requirement is for the subject to understand and identify them consistently.

The physiological signals measured and analyzed in [1] are listed in Table 2.3 along with the respective sensor placement. From the sensor orientation, one can observe that the aim was to stabilize the left hand while leaving some freedom of movement to the right hand. The experiment lasted for 30 days and every day the subject was recorded

for 25 minutes. Emotion elicitation was based on imagery exercises. The volunteer was asked to pick scenes or situations from her own experience that could induce the desired emotion with adequate valence.

Picard *et al.* [1] suggested for classification features picturing the statistical properties of the signals rather than just their changing rate. These statistical features are listed in Table 2.4. In addition, and in order to account for the effects of non emotional factors, an extra set of physiology related features was calculated. These features along with a description can be found in Appendix C.

Having acquired a feature set, the next step of the training procedure is to *select* the appropriate ones and *transform* them. The authors described a combination of the Sequential Floating Forward Search algorithm (SFFS) in conjunction with Fisher Projection (FP) [62]. SFFS is a statistical searching methodology, which performs an nonexhaustive search by continuously adding and removing features from a set by checking their classification power. FP is a method for dimensionality reduction, which can lower the dimensions of the feature space down to $C - 1$, where C is the number of classes, i.e., emotions the system can identify. These two methods were combined hierarchically in one framework, by using the output of SFFS as input to FP. For classification, a k-Nearest Neighbor scheme was employed with k varying between 1 and 20. The highest classification performance reported in [1] was 83.3 % for the identification of three emotions.

Scheirer *et al.* [2], reported a novel experimental procedure to induce and measure frustration. This new approach had considerable advantages to the synchronization problem. More precisely, instead of deploying imagery or audio visual induction techniques, they have created a computer game which aimed to frustrate the user. At random interval of times the game failed (mouse clicking did not work) this way spoiling the players' pleasure, who got eventually frustrated [63].

Utilizing a computer game to provoke emotion has significant advantages for signal

Mean	$\mu_x = \frac{1}{N}\sum_{n=1}^{N} X_n$
Standard Deviation	$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \mu_x)^2}$
Mean of first differences or raw samples	$\delta_x = \frac{1}{N-1}\sum_{n=1}^{N-1}\lvert X_{n+1} - X_n\rvert$
Mean of first differences or normalized samples	$\widetilde{\delta}_x = \frac{\delta_x}{\sigma_x}$
Mean of second differences or raw samples	$\gamma_x = \frac{1}{N-2}\sum_{n=1}^{N-2}\lvert X_{n+2} - X_n\rvert$
Mean of second differences or normalized samples	$\widetilde{\gamma}_x = \frac{\gamma_x}{\sigma_x}$

Table 2.4: Typical classification features used in [1].

labeling. In the affective computing research, one of the major challenges is generating accurate datasets, and this is first because it is extremely difficult to elicit emotions and second because the ground truth is not always well defined. To this end, the ground truth represents the true labeling of the signals with respect to the onset of an emotional phase. What Scheirer *et al.* suggested in [2], was that using a computer game to frustrate the user, the time instances of frustration could be easily labeled. However, that alone is not very accurate and latency periods should be taken into account as there is usually a short time interval before a human actually experiences a feeling internally (three seconds for galvanic skin response [2]).

Scheirer *et al.*'s work [2] although focusing on human frustration could have been deployed for any other emotion by accordingly changing the computer game. In [2], frustration is defined as a manifestation of negative arousal when something impedes the subject's progress toward a goal. To quantify this arousal and distinguish between frustration and non-frustration states, two physiological signals were considered, the galvanic skin response and blood volume pressure. Once more, this selection of signals was not compulsory and the authors explain that this choice might not even be optimal. However, for the purpose of their experimental setup, GSR and BVP were suitable because

they can be measured unobtrusively by occupying only one of the subject's hands. For the analysis in [2], data from 24 subjects were utilized. The data were modeled with a Hidden Markov Model (HMM) because of its success in describing dynamic systems like speech recognition. A set of features was proposed for the discrimination of frustration based on the variability of GSR and BVP. More details pertaining to these features can be found in Appendix C. During training, the highest classification performance achieved was 81.87%, which dropped to 67.4% for the testing set. Given that only two classes were considered this result is slightly better than random guessing. The major novelty of [2] is in setting the standards for the experimental procedures in the affective computing research. The authors explained that the poor performance reported, could be addressed with more sophisticated pattern recognition and machine learning techniques.

In 2003, Nasoz *et al.* [64], introduced the idea of adapting a software interface according to a user's current emotional state. A multimodal framework was proposed, based on the combination of facial, verbal, gesture and physiological recordings. A synthesizer was designed to combine all this information and provide the user with feedback.

The emotional states explored were six: sadness, anger, surprise, fear, frustration and amusement. They were elicited using visual stimuli, movies particularly chosen to relate to the above mentioned psychological states. The biosignals measured were the galvanic skin response, heart rate and body temperature. Data from 31 subjects were divided into training and testing and two algorithms were used for classification, namely the k-NN and and the discriminant function analysis. (DFA). The two algorithms were tested independently to show that DFA outperforms k-NN with a maximum recognition performance of 90% for fear.

A later and more detailed report by the same authors [65] placed the emphasis to presence technologies. The MAUI (Multimodal Affect User Interface) prototype system was presented to enhance social presence using physiological signals. Facial expression recognition was also central to MAUI, as one of its main features was an avatar who could

mimic the user's expressions. This is because when characters (personas) are employed in virtual environments, their ability to communicate emotions increases significantly the social presence and co-presence.

The list of emotions considered was the same as in [64]. Their study involved 14 subjects who were asked to watch movie clips from a predefined list of films that had been shown to elicit specific emotions successfully [66]. Physiological signals (galvanic skin response, heart rate and temperature) were measured non invasively with a wireless armband. An important step in feature processing was normalization of the measurements with respect to the respective relaxation value. This was done in order to avoid the individual differences of participants. The main analysis in [64] included the classification methods explored previously; k-NN, DFA plus a newone, namely the Marquardt Backpropagation (MBP). The MBP method was reported to outperform the rest in recognizing all emotions (for instance 92% for sadness).

In 2004, Kim *et al.* [67] reported large scale experiments for physiological emotion recognition with more than 100 subjects. The methodology relied on features out of three physiological measurements i.e., skin temperature, heart rate and galvanic skin response. A strong justification of the relationship between these signals and the autonomic nervous system's reaction to stimuli can be found in [67]. The analysis was based on two experiments; the first involved 125 subjects (5-8 years old) under a variety of audio and visual stimuli. The second experiment was carried out a year later to record 50 more subjects (7-8 years old).

The heart rate variability (HRV) information was obtained through ECG with peak detection (using the Teager energy operator) and calculation of the R-R intervals. The authors suggested the spectrum of HRV as a feature source. Given that the spectrum estimation should be accurate even for a short term signal (so that real-time approximations can be obtained), time-series models for frequency estimation were used [68]. More specific, two sub-bands were considered, a low frequency one at 0.03-0.15 Hz and

a high frequency at 0.15-0.4 Hz. For the GSR, feature extraction comprised detection of response occurrences which are waves of various amplitudes and lengths depending on the stimuli. To do that, the GSR signals were differentiated and smoothened in order to apply a threshold for the detection of the waves' onsets and offsets . Four features were used for classification out of GSR i.e., the mean DC level, mean values of GSR detected wave amplitudes and duration, and the number of event (interesting wave) occurrences in a 50 second segment. Finally, the skin temperature was used as a feature without special processing.

The signals were divided into training and testing sets for classification among three or four discrete emotional states i.e., sadness, anger, stress (and optionally surprise). The recognition performance for three clusters of data from the first experiment was 55.2%, and 78.4% for the second experiment. According to [67] the later dataset included signals of higher quality. Among four emotional states, the correct classification rate of the second dataset was reported to be 61.8%.

In 2005, Healey *et al.* [3] suggested the use of physiological signals for stress detection under real world driving tasks. The purpose of the proposed system was to analyze a driver's psychological status in order to provide feedback or take appropriate action. An adaptive framework was envisioned, where the computer would cope with the driver's stress by diverting cell phone calls to voice mail, suggest appropriate music and reduce workload by manipulating the navigation systems.

Four signals were considered in [3], ECG, EMG, GSR and respiration (R). Among the challenges of that implementation was for the system to react to stress detection in real time. However, none of the monitored biosignals responds in time for appropriate adjustment of vehicle control (1-3 min lag) [3]. To that end, two experimental setups were considered:

1. *Analysis I.* The first analysis attempted to recognize general stress levels (defined as low, medium and high), by utilizing data of 5 minutes long. In addition, this

module was designed to detect stress based on a systematic fusion of all four signals using standard pattern recognition techniques. Overall, 21 features were estimated and used for classification, listed in Table C.3 in Appendix C.

2. *Analysis II.* The second analysis aimed at a more detailed examination of the physiological signals. For this reason, the signals were processed independently, and at a shorter time span i.e., every second. A special continuous stress metric was designed to judge the environmental stressors at every second of the experiment, in order to investigate the correspondence between this measure and the biosignals. The stress metric was computed in a way that it revealed the instantaneous workload of the driver, according to videos taken throughout the drives.

Classification on features from the first analysis indicated high accuracy, with low level stress being correctly recognized for all testing data (100%), medium and high stress recognition rates were 94.7% and 97.4% respectively. For Analysis II, continuous time features were designed out of each of the biosignals, in the same range as the stress metric, for the correlation coefficient (in the range $[-1, 1]$) to be used for matching. The highest performance was achieved for a GSR metric picturing the instantaneous variation of the mean, with the correlation coefficient reaching 0.99.

Driving applications of affective computing were revisited later by Hönig *et al.* [69]. By recording 24 subjects in a driving simulation environment, the authors attempted to recognize the various stress levels provoked by driving related tasks such as changing lanes, along with memory task such as question answering and arithmetic. Six physiological signals were monitored during the simulation: ECG, EMG, GSR, BVP, temperature and respiration along with some of their derivatives (ECG or BVP heart rate, reaspiration rate and so on).

During the simulations, the subjects' facial expressions were recorded with a camera to be used for evaluation of stress elicitation. Extra subjects were used as labelers i.e.,

to watch the video recordings and assess the stress levels based on the facial reaction of the volunteers. This labeling, along with the standard stress onset timing designation provided by the system (for instance, the instance when the driver is required to change lane), were used in order to establish the data ground truth.

To cover a range of short to long period reactions, feature extraction operated on signals of length varying between 1 and 60 seconds. A large number of features was computed in [69] based on the mean, standard deviation and slope of the various windows. The Fisher linear discriminant analysis was used for dimensionality reduction and classification was carried out based on Gaussian mixture model probabilities. Overall, and by fusing information from all physiological signals, a correct stress recognition rate of 88.8% was achieved.

Lee *et al.* [70], suggested the use of a neural network that would be trained on emotional cues and would then be tested on data not previously seen. Interestingly, only 6 subjects participated in the experimental set up. According to the authors that was because the number of subjects involved is independent from the overall performance as emotions are experienced in a different physiological manner by every individual. Two physiological signals were considered: the galvanic skin response and the heart rate, which was acquired with QRS peak detection from the electrocardiogram. The heart rate was processed in the frequency domain, by estimating the ratio between low and high frequency coefficients of the spectrum.

A multilayer perceptron (MLP) neural network was trained for the recognition of four emotional states: sadness, calm pleasure, interesting pleasure and fear. All emotions were induced by video clips, that the volunteers had chosen themselves. After watching each video, the subjects were asked to describe their emotional experience using the self assessment manikin (SAM). The overall performance for all emotional clusters was 80.2%, with fear being the best physiologically distinguishable feeling.

Anttonen *et al.* [71], focused in designing technologies which are not only unobtrusive

to the user but also offer no distractions. Traditional equipment for physiological signal measurements are wired sensors which need to be attached to the human body by trained personnel. This limits significantly the range of applications as mobility is restricted to a small area. In addition, such sensors fail a lot because they get easily detached, and distract the subject who is expected to focus in experiencing an affective state. For this reason, the authors in [71] suggested the use of a chair which is equipped with electromechanical film (EMFi chair) to be able to record physiological measurements while someone is sitting on it.

An initial step in Anttonen *et al.*'s [71] experimental setup was to ensure that the chair can collect sufficiently accurate data. Therefore, they used a simple photoplethysmography (PPG) sensor attached to the earlobe of the subject while he/she was sitting on the chair. Both the chair and the PPG sensor offered heart rate measurements, which were compared with each other to demonstrate consistency in the chair recordings (the correlation between the two was 0.99). This allowed the authors to rely on the chair recordings for the rest of their experiments.

The objective of the main experimental procedure in [71] was to classify among positive, neutral and negative emotional states. 26 subjects were recorded sitting on the EMFi chair, while their psychological state was stimulated via audio and visual stimuli (International Affective Picture System (IAPS) [72] and International Affective Digitized Sounds (IADS)). Both positive and negative stimuli resulted in heart rate deceleration compared to the neutral reaction. What is more, the deceleration under negative effects was greater than the positive in magnitude and in length. Heart rate changes in positive stimuli were observed only during the first two seconds and right before returning to a normal rhythm. The latency period for the manifestation of heart rate changes under negative reactions was approximately six seconds from the onset of the stimuli.

Nakasone *et al.* [73], presented a framework for machine emotion recognition which concentrated on the implementation challenges of this technology. More precisely, the

authors looked at gaming scenarios between humans and humanoid agents, who were envisioned to "read" the player's psychological state and respond appropriately. The system was designed upon five main layers namely, the synchronization, the device, signal categorization, bayesian network and interface layers.

The synchronization layer was set responsible for the initialization of emotion assessment and communication among the rest of the layers. For instance, data acquisition, storage and analysis where synchronized in this module. The device layer controlled the physiological recording devices, which measured EMG and GSR. At the categorization layer, the signals were subjected to feature extraction, which involved a simple calculation of their sample mean and subtraction from a baseline value (where the baseline was considered to be the physiological response at rest). The bayesian network layer would utilize the classification results in order to finalize the emotion decision and finally, the interface layer adapted the game to the particular player status.

Although no specific experiments were carried out, Nakasone *et al.* [73], addressed problems related to the actual realization of affective computing. Overall, three challenges were outlined: 1) The baseline problem, which relates to the standard resting conditions against which instantaneous physiological measurements should be compared, 2) The timing problem, i.e., detecting the exact onset of an emotional phase and 3) The intensity of a particular experienced emotion. To make sure baseline signals were obtained for comparison, the authors suggested an initial relaxation period, where the subject would listen to calm music. Furthermore, to ensure that emotions were captured in the signals, the authors suggested sampling every 50 msec.

In 2008, Benovoy *et al.* [74], proposed another novel application for affective recognition systems. The main idea was for emotion assessment to be used in a performance setting, for instance a musical act, were the performer leads an instrument to compose an artistic output based on the instantaneous physiological reaction of a subject. This application had two main design phases:

1. Biosignal recording and mapping with emotional states. The blood volume pressure (BVP), galvanic skin response (GSR), electromyography (EMG) and respiration were considered for classification among four emotions (joy, anger, sadness and pleasure).

2. The second stage was meant to provide a rich external manifestation of a person's internal emotional status [74]. The purpose was to build a musical device which would be mapped to the user's feelings but also controlled by him/her.

Given the specifics of the application, a direct interaction with the experimental subject was preferred over standard emotion induction methodologies used previously (audio visual). A professional actor guided by an experienced theatre director was monitored while experiencing a variety of characters and situations, particularly designed to elicit the desired emotions. By the end of each experiment the actor was asked to fill in a questionnaire concerning the intensity of the previously experienced emotion.

A large variety of features was estimated from the recorded biosignals. Overall 225 features were used, similar to [1]. A list of these features can be found in Appendix C. The next step was to select the optimal for classification features using the sequential forward selection (SFS) algorithm. The overall dimensionality was reduced with Fisher discriminant analysis, and finally classification was carried out with either the linear discriminant analysis (LDA), or k-NN or a multilayer perceptron (MLP) neural network. It was demonstrated that the LDA outperforms the rest of the classification schemes, with a classification rate of 90% for four emotions. This notable result was attributed to the fact that a professional actor was recorded, thus emotion elicitation was more successful compared to prior trials.

In summary, prior DEM approaches have established biosignals as a carriers of affective information. The simplicity in the conception of emotion within such models (i.e., the discretization to distinct emotional conditions) makes them easily employable in real

life systems. However, prior works require the collection of more than one physiological measurements in order to assess emotion, which limits real life applications due to increased obtrusiveness. In addition, the ECG signal has been primarily used for HR estimation while the evolution of the waveform under emotional activity has not been studied.

2.4.2 Affective Dimensional Models

In 2004, Mandryk *et al.* [75], established the applicability of physiological emotion recognition in the evaluation of entertainment technologies. The breakthrough of this work lies in proving that there is a different physiological response when users play against other players, as opposed to playing against the computer. As a matter of fact, there is a different reaction when playing against friends compared to strangers [76]. In addition, it was shown that people tend to immerse and enjoy a game more, when they play against humans [75]. These observations and the respective technology apply directly to the entertainment industry for the design of virtual environments and computer games.

Mandryk outlined the three challenges of this research in [77]:

- *What is the definition for a successful game?* A game can not be characterized successful based on the user performance, but on the user's affective experience [77]. A prosperous game environment enhances interaction between the players (co-presence), and only this can offer a truly immersive user experience. Thus, success is measured in terms of the psychological engagement of the players.

- *How can the entertainment variables be measured?* A means to quantify user engagement, is by measuring his/her emotional involvement. The interest is not in defining *which* emotion is experienced, but the level of subject's arousal in general. The biosignals considered for this purpose are the galvanic skin response, various cardiovascular measurements (electrocardiogram, heart rate, blood pressure),

respiration rate and amplitude, and electromyography.

- *How can a measurement be associated to the player's experience?* The game events were monitored to label the signal recordings. This approach is similar to the Scheirer *et al.*'s [2] idea for controlled emotion elicitation which allows for better signal segmentation.

The features used for discrimination of arousal were simply the time average of the signals. This kind of features diminished the variability of each psychological state, but this was intentional since the authors were interested only in detecting the presence of emotions and not in defining which emotion was experienced. The experimental set up involved 10 male subjects and it was shown that the arousal was greater for all physiological measurements when playing against humans.

In 2007, the same researchers [78] managed to design a consistent mapping between physiological measurements and emotional states i.e., the arousal-valence (AV) plane. This time the authors were interested in identifying the exact emotional state (fun, excitement, frustration, challenge and boredom). A fuzzy approach was adopted to systematically combine information from four physiological measurements (galvanic skin response, heart rate, smiling and frowning electromyograph) and then project it to meaningful valence and arousal dimensions. Accordingly, a second fuzzy scheme established the correspondence between the AV axes and the desired emotional states.

The four biosignals served as inputs to the first fuzzy framework. A preprocessing step was required to ensure that the measurements were comparable with each other (this was required by the fuzzy model). Preprocessing involved smoothening with a moving average window and then normalization into a percentage between 0 and 100.

The experimental setup in [78] included 12 subjects. The physiological readings from the first 6 were used for training i.e., to analyze emotional response and design the fuzzy logic, and the other 6 subject recordings were used for testing. To evaluate the performance of the algorithm, the *subjective* reports (questionnaires) were compared

against the *objective* recognition results (from the fuzzy scheme) to reveal similar trends for most emotional classes. Even though the performance of the framework was not presented analytically, Mandryk *et al.* were the first to attempt a direct match between physiological signals and the AV plane. The reader should note that prior to this work, researchers have concentrated in modeling specific emotional states.

Haag *et al.* [79] proposed an affective dimensional model that can classify information on an arousal and valence basis. A collection of physiological signals was used for this purpose, each processed independently and according to its special characteristics. The biosignals were the electromyogram (EMG), galvanic skin response (GSR), skin temperature, blood volume pulse (BVP), electrocardiogram (ECG) and respiration.

The authors in [79] comment strongly on the importance of the emotion induction methodologies and to the effects on the performance. To elicit emotions appropriately, strong realism is required which is usually not in accordance with the experimental ethics (provoking severe negative emotions is ethically questionable). For this reason they used a collection of pictures from the International Affective Picture System (IAPS) [72]. The pictures were classified into six sets, each consisting of 5 photos. Every set was labeled as low, medium or high valence, or low, medium or high arousal. The combination of the two extensions (arousal and valence) defined a two dimensional space for the classification of the physiological features.

The statistical features employed for classification were:

- The standard deviation of the heart rate, estimated with peak detection out of the ECG signal.
- The standard deviation of the heart rate, estimated with peak detection out of the BVP signal.
- Standard deviation of BVP amplitudes.
- The EMG amplitude.

- Standard deviation of skin conductivity (GSR).
- Standard deviation of respiration amplitude.
- Standard deviation of respiration rate.

It should be mentioned that the above mentioned list of features was the optimal selection, and more features were tested and rejected in [79] because of poor performance. One subject was tested on different days and also various times of the day. The recognition rate was reported for correct classification considered within 10% and 20% distance from the true class on the AV plane. The second option increases the bandwidth thus greater performance was expected. More specific, for the large bandwidth a recognition rate of 96.58% and 89.93% was achieved for arousal and valence respectively.

Jones *et al.* [80], proposed a system for recognition of affective states based on dimensional modeling. The main idea is that an emotional state can be sufficiently characterized by its arousal and valence magnitudes, without mapping them to distinct emotions (sadness, happiness etc). The two dimensional AV model was used as a basis for projection of information collected from physiological signals.

The plane was divided into 25 areas, each representing an emotional region. Three physiological signals are considered for recognition, the blood volume pressure (BVP), skin conductance (GSR) and the respiration rate. Emotions were induced with 21 images from IAPS [72] since these images have already been rated in valence and arousal by psychologists. Each picture was displayed to the subject for 15 seconds after which there was an interval of rest for 25 seconds.

As with any pattern recognition problem, collection was followed by feature extraction and classification. The features used in [80] were simple statistical measures of the signals like the mean and variance. For classification, two multi layer perceptron neural networks were used, one for each descriptive dimension (AV). Overall, 13 subjects were recorded. The data of 10 subjects were used for training and the rest for testing. The authors

		Ekman et al.	Sinha et. al	Picard et al.	Scheirer et al.	Nasoz et al.	Kim et al
Emotional Model	**Type**	DEM	DEM	DEM	DEM	DEM	DEM
	Emotions Induced	Surprise, Disgust, Sadness, Anger, Fear, Happiness	Fear, Anger, Joy, Sadness	Anger, hate, Grief, Platonic love, Romantic love, Joy, Reverence	Frustration	Sadness, Anger, Surprise,Fear, Frustration, Amusement	Sadness, Anger, Stress, Surprise
	Elicitation Method	Imagery Task	Imagery Task	Imagery Task	Computer Game	Movie clips	Audio, Visual stimuli
Biosignal Processing	**Signals**	HR, T, GSR, Muscle tension	HR, BVP, and other cardiovascular measures	EMG, BVP, GSR, R, HR	GSR, BVP	GSR, HR, T	T, HR, GSR
	Feature Extraction	Change Scores	Change Scores	Mean and std analysis with SFFS and Fisher projection.	Variance and pinch measures.	Signal Average for each emotion.	HR spectrum, GSR wave features
	Classification	Decision Tree	MANOVA	k-NN	HMM	k-NN, MBP, discriminant analysis.	SVM
Performance	**Number of subjects**	16	27	1	24	14	50
	Classification Rate	Not reported	Not reported	83.3% for 3 emotions	67.4%	92% for sadness	61.8%
Comments		Professional actors were recorded.	Fear and anger can be differentiated through the diastolic blood pressure.	Evaluation of subject dependent method. Experiments with the same subject over several days.	The experimental process was controlled with the use of a specially designed computer game.	Emphasized the use of affective computing in presence technologies.	Special properties of the biosignals were considered for appropriate feature extraction.

Table Notation

DEM Discrete Emotional Model **ADM** Affective Dimensional Model **HR** Heart Rate **T** Temperature

GSR Galvanic Skin Response **BVP** Blood Volume Pressure **R** Respiration **ECG** Electrocardiogram

EMG Electromyogram

Table 2.5: Affective computing using biosignals: Comparison Milestones

		Lee et al	Anttonen et al.	Benovoy et al.	Mandryk et al.	Haag et al.	Jones et al.
Emotional Model	**Type**	DEM	DEM	DEM	ADM	ADM	ADM
	Emotions Induced	Sadness, Calm Pleasure, Interesting pleasure, fear	Positive, neutral and negative status	Joy, Anger, Sadness, Pleasure	AV plane	AV plane	AV plane
	Elicitation Method	Video clips	Audio, Visual stimuli	Imagery Task	Computer games against a friend or the computer	Visual stimuli	Visual stimuli
Biosignal Processing	**Signals**	HR, GSR	HR	BVP, GSR, EMG, R	GSR, HR, smile EMG, frown EMG	EMG, GSR, R, T, BVP, ECG	BVP, GSR, R
	Feature Extraction	Signals used for classification directly	ANOVA	Mean and variance features selected with SFS and Fisher discriminant	Signal histogram analysis	Standard deviation features	Mean and variance features
	Classification	MLP Neural Network	-	k-NN, MLP, LDA	Fuzzy logic Method	Neural Network	Two MLP neural networks for Arousal and Valence.
Performance	**Number of subjects**	6	26	1	12	1	13
	Classification Rate	80.2%	62.5%	90%	*Arousal 3% Valence 6%	96.58%	62% Valence 67% Arousal
Comments			Experiments with unobstrusive sensors.	Emphasized the use of affective computing in artistic performance settings.	Defined the standards for the application of affective computing in the entertainment industry		

* Performance measured in terms of distance between the AV values obtained with the fuzzy scheme and the subject reports.

Table 2.6: Affective computing using biosignals: Comparison Milestones (continued)

		Healey et al.	Nakasone et al.	Honig et al.
Emotional Model	**Type**	DEM	DEM	DEM
	Emotions Induced	Low, Medium, High Stress	Relaxation, Joy, Frustration	Relaxation, Stress
	Elicitation Method	Driving Task	Computer card game	Driving simulations combined with oral stressors
Biosignal Processing	**Signals**	ECG, EMG, R, GSR	EMG, GSR	ECG, EMG, GSR, R, T, BVP
	Feature Extraction	Statistical and Spectral features	Signal mean	Mean, Standard deviation, Slope and LDA
	Classification	Discriminant Analysis	Bayesian network	GMM
Performance	**Number of subjects**	24	No experiments reported	24
	Classification Rate	100 % Low stress 94.7% Medium stress 97.4% High stress	No experiments reported	88.8 %
Comments		Performed experiments for real time stress detection in real life scenarios.	Proposed a layer design for real affective computing applications in gaming scenarios.	The biosignals were labeled with special attention leading to impressive stress detection rates.

Table 2.7: Affective computing using biosignals: Comparison Milestones(continued2)

described two performance measures, based on the location of the system's output on the AV plane. The first measure counted as correct classification cases only outputs which conformed exactly with the true class on the AV plane. With this arrangement, the performance was 30% and 35% for the valence and arousal dimensions respectively. The second performance measure was less strict and considered successful a classification within +/- 1 unit of the true class. With this arrangement, the system achieved 62% and 67% correct classification for the valence and arousal dimensions respectively.

In conclusion, prior works in this field have managed to establish the idea of affective computing along with its connection to psychophysiology. Various biosignals have been examined and used for recognition based either on discrete or dimensional modeling of emotions. Affective computing is a multi-disciplinary field inviting works from the computer science, statistics, psychology, engineering, social and medical fields. Even though the emotional intelligence problem with biosignals has been studied for years, no affective modeling methodologies has been proposed to perform pattern recognition using the signals' inherent properties.

Tables 2.5 2.6 and 2.7 outline some comparison milestones among previous works in the field. It can be inferred from these Tables, that even though the experimental procedures and emotion modeling have been advancing throughout the years, there are limited attempts to design affective pattern recognition techniques, and feature extraction usually operated on standard statistical measurements (mean, standard deviation etc.).

2.5 Chapter Summary

This chapter presented a discussion of the ECG signal from two different perspectives i.e., from a physiological (anatomical) and psychological point of view. These interact on the ECG signal in a non-linear fashion.

The essence of biometric recognition is the detection of physiological patterns that are robust to psychological variations. Similarly, the observation of psychological cues, irrespective of subject-specificity, is the foundation of affective computing. Therefore, biometric recognition and affective computing have contradicting interests on the ECG signal.

To ensure strong biometric templates, one needs to guarantee robustness to psychological variations that may affect the signal. Ideally, we need to isolate the physiological aspect of the ECG. However this may not be mathematically feasible given the complex nature of the signal. Moreover, psychology itself is a very complex science wherein one needs to define the exact conditions that may threaten the biometric template.

The first step in this thesis is to present the problem of biometric recognition, with psychological factors not taken into account. The objective of this analysis is to demonstrate the perils of ignoring time dependency from an identity verification point of view.

The next step is to examine the signal from a psychological point of view. While prior works in the affective computing literature have marginally used the ECG waveform, this thesis argues that under particular conditions this signal can exhibit emotion specific patterns which are associated with instances of biometric template destabilization. Finally, a solution to this problem is presented to accommodate biometric recognition in monitoring settings wherein template destabilization is most prominent.

Chapter 3

ECG Databases

3.1 Experimental Protocols

This chapter presents the experimental protocols that have been designed for the collection of ECG signals. The data from the subsequent experiments have been organized, anonymized and securely stored. In addition, all databases have been made publicly available, by the Biometric Security Laboratory, and can be provided upon request.

When evaluating performance on real ECG signals, it is very important to establish experimental procedures which ensure that the necessary information is captured and in the data. Although a number of ECG databases are publicly available, they are not suitable for biometric testing. Since the signal is typically used in medical diagnostics, existing databases offer ECG signals which depict a number of cardiac conditions.

For biometric testing the database needs are different than for clinical testing. First of all, the subjects are required to be healthy or at least to not exhibit prominent cardiac irregularities. Although it is of interest for the biometric system to tolerate small irregularities, the ECG signal needs to be first established as a biometric modality for healthy individuals. Furthermore, in addition to the *uniqueness* evaluation of the ECG, one needs to also examine the *permanence* of the signal with time. A typical way to do

this in the biometrics field is to collect the ECG signal from the same individual during different recording sessions.

For the above reasons, two different experimental setups have been designed for biometric testing. In the first experiment (*Short-term recording experiments*), the objective was to collect ECG signals from as many subjects as possible (to evaluate uniqueness), during two different recording sessions (in order to evaluate permanence). This experiment is referred to as *short-term* because the collection periods in every session were very short. In the second experiment (*Long-term recording experiment*), longer ECG readings were acquired for every individual.

For the purpose of affective computing, this chapter also presents experimental procedures for ECG signal collection under emotional activity. These experimental procedures are more complex than typical biometric setups, because one needs to guarantee successful emotion elicitation. This chapter presents two such experiments i.e., a *passive* and an *active arousal induction experiments*. The difference between the two is that in the first the subjects were passively exposed to an emotional stimulus, while for the second case the subjects participated actively in the experimental procedure.

3.2 Short-Term Recording Experiments

The short-term recording experiments took place at the Biometrics Security Laboratory of the University of Toronto (Ethics protocol # 23018). 52 volunteers participated in this experiment, the majority of which were graduate students. Overall, the age range was between 22 and 50. For this experimental setup, two recording sessions took place several weeks apart, in order to investigate the permanence of the signal in terms of identification performance. All participants were asked to repeat the experiment, however only 16 were recorded at both sessions.

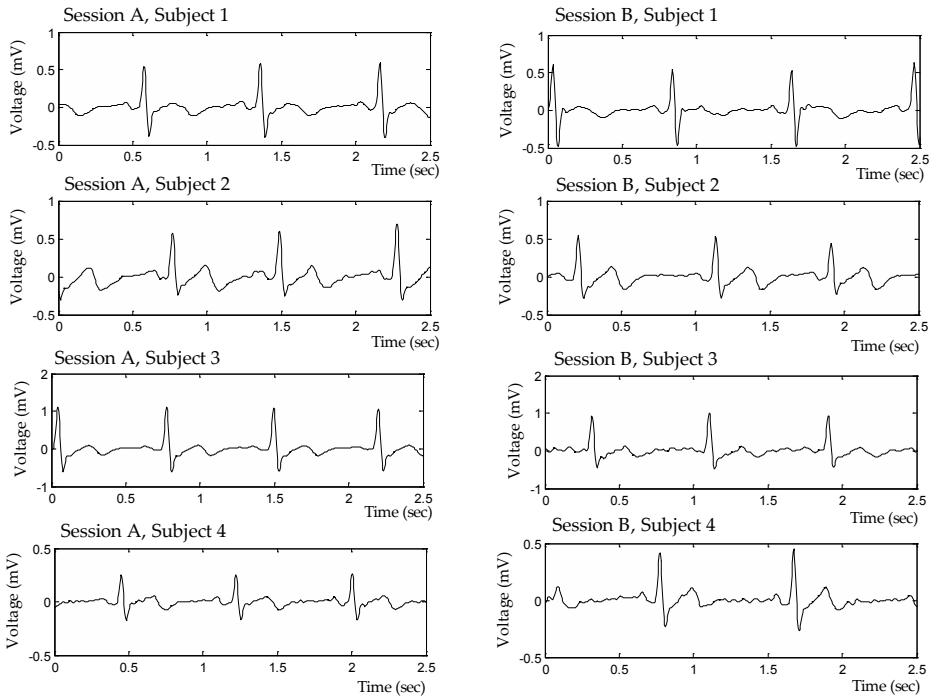


Figure 3.1: ECG samples from the short-term recordings database. Every subject was recorded during two sessions.

For the short-term recording experiments, a Vernier[1] ECG sensor (EKG-BTA) with the Go!Link connector were used. Lead I ECG was collected, by placing the electrodes at the inside of the right and left elbows, and the ground on the right wrist. The sampling frequency was set to the device's maximum (200Hz) and the recording duration to 3 minutes. The experimental procedure was disclosed to the volunteers in the beginning of the experiment, who also signed consent forms. During the collection, the subjects were given no special instructions, in order to allow for mental state variability to be captured in the data. Figure 3.1 shows some examples of ECG signals from this database, from four different subjects.

In order to evaluate the performance of a biometric algorithm, the earliest recording are typically used for enrollment of the user in the system and the latter for testing.

[1]http://www.vernier.com

For the 36 individuals that only one recording is available, different treatment can be performed. For instance, when a generic pool is necessary, the ECG signals from these individuals are used to form it.

3.3 Long-Term Recording Experiments

The purpose of the long-term recording experiment was to explore the dynamic nature of the ECG signal and to investigate in greater detail its instantaneous changes. For this experiment, all subjects were debriefed at the Biometrics Security Laboratory, however the collection was performed in their personal working stations. 10 volunteers participated in this experiment. The recording sessions were approximately 2 hours for every individual.

The Hidalgo Equivital EQ-01 monitor[2] was used for the collection. This device allows for long term wireless ECG monitoring. It was chosen for this experiment because unobtrusive data collection was necessary. The signals were recorded from the chest area, and wirelessly transmitted to a computer. In order to account for the particular emotional changes, no special instructions were given to the participants. After connection establishment and testing of the equipment, subjects signed consent forms and returned to their personal working stations, to continue their daily activities for 2 hours. ECGs were digitized at 256 Hz.

3.4 Passive Arousal Experiment

The passive arousal experiment was conducted at the Affect and Cognition Laboratory, of the University of Toronto (Ethics protocol # 22059). The purpose of this experiment was to evaluate the response of the ECG signal to passive emotional stimuli.

[2]http://www.equivital.co.uk

The International Affective Picture System (IAPS) [72] was used in this experiment as a passive stimulus for emotion elicitation. This picture set provides ratings of affect for a large set of emotional conditions using a wide range of semantic categories. The pictures have been rated by more than 100 subjects at the University of Florida based on three criteria: arousal, valence and dominance. Efforts have been made to include pictures that cover the range of these norms [72]. The IAPS photo-set is widely deployed in affect research as a passive emotional stimulus.

In this experiment, the volunteers were asked to sit conveniently and inspect images of the IAPS photoset, presented to them as a slide-show. The images attempted to induce the following emotional conditions: *gore, fear, disgust, excitement, erotica* and *neutral*. These conditions cover the valence spectrum from negative (gore, fear, disgust) to positive (erotica, excitement). Also, with respect to the arousal, the gore and erotica stimuli are considered to have induced the most highly arousing states (see Appendix B).

Five pictures of the same emotional target (emotion batch) were displayed in random order. The continuous display of pictures with similar valence and arousal attempted to create an "emotional state" rather than an instantaneous emotional reaction, in order to account for potential cardiac latencies. Every batch was repeated twice with a different set of pictures. Between batches the subject was asked to perform a simple arithmetic task that would bring the psychological activity to baseline. Lead I ECG was collected during the slide-show, at 1KHz, with the BIOPAC MP 150 system. Upon completion of the experiment, the volunteers were instructed to look at the pictures again, and report subjectively valence and arousal.

In total, 44 volunteers, of ages between 21-40 participated in this experiment. Recordings from 12 individuals were discarded due to noise and artifacts. The BIOPAC system allowed for automatic synchronization of the ECG signal with the pictures. Every emotional state is assigned with a unique number. A vector of the same length as the signal, such as the one shown in Figure 3.2, was provided by the system, where these unique

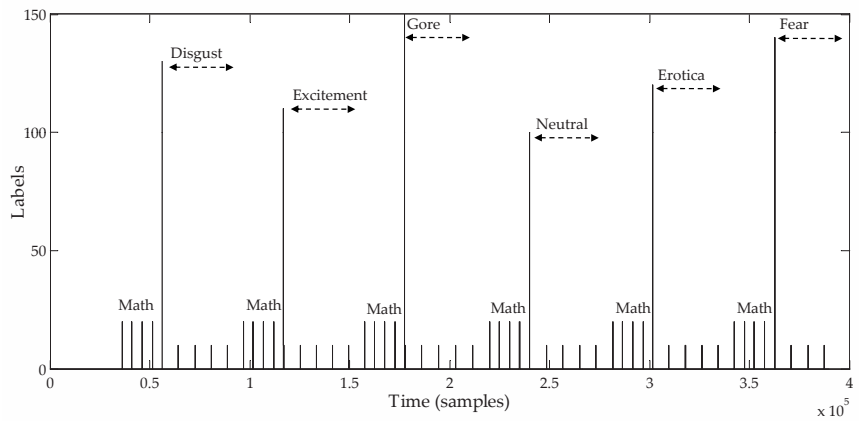


Figure 3.2: Data labeling for the passive arousal experiment. Every picture of the IAPS photo-set is assigned to a unique number which indicates the beginning of the respective emotion on the data.

Figure 3.3: Game and face video playback, used for self-assessment of arousal.

numbers indicated the onset of the respective emotional state

3.5 Active Arousal Experiment

The active arousal experiment was also conducted at the Affect and Cognition Laboratory, of the University of Toronto (Ethics protocol # 25419). In this experiment, a commercial video game was used to elicit active mental arousal.

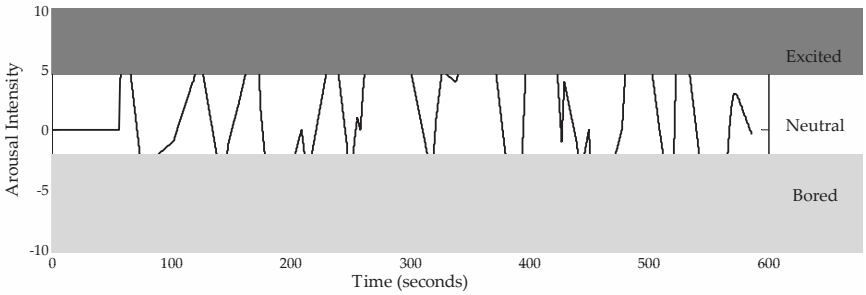


Figure 3.4: Data labeling for the active arousal experiment. The FEELTRACE is a continuous arousal indication.

The game was designed to present increasing difficulty. The goal was to have the player gradually immersed, by increasingly concentrating in order to meet the game requirements. A popular video game was used, namely the Cube 2: Sauerbraten[3], developed in 2009. This game allows the user to dynamically pre-edit the environment and control the difficulty levels. A pilot game was built on this ground, to assist the needs of the experiment. The subjects got motivated with deception, by letting them know that the purpose of the experiment is to measure game completion time.

All participants were seated in front of a computer screen, and presented with a short introduction to the video game. A five minute pilot game will was played, for the participant to learn and be adjusted to the game. During that time, no physiological response was monitored. When the subject felt comfortable with the process, the ECG sensor was placed and the main game was initiated. In total 43 volunteers participated in this study. Data from one person were discarded due to noise.

Depending on the familiarity of the subject with game playing, the duration of the experiment varied between 20-45 minutes. During the game, ECG was monitored using Hidalgo's Equivital sensor, which is portable and wireless. Unobtrusiveness was very important for the subjects to naturally immerse to the game. ECG was recorded from the chest, and digitized at 256Hz. Because of the chaotic nature of the game, and the

[3]http://sauerbraten.org

unforeseeable order of events that can take place, arousal annotation was self-determined. For this reason, a video of the player's facial expressions was captured during the game (synchronous to ECG). Upon game completion, the subjects were asked to watch a playback video of the game and their facial expressions while continuously reporting arousal using FEELTRACE [81]. Figure 3.3 shows an example of such a self-assessment video.

In order to compare the effects of active and passive arousal, we asked the same volunteers to participate in both experiments, to acquire subject specific arousal data for both scenarios. In total, signals from 31 volunteers were common and eligible for processing.

	Sort-Term ECG Database	**Long-Term ECG Database**	**Passive Arousal Database**	**Active Arousal Database**
No of Subjects	52	10	44	43
Rec. Duration	3 minutes	2 hours	15 min	20-45 min
Sensor Placement	Lead I	Chest	Lead I	Chest
Rec. Device	Vernier EKG-BTA	Equivital EQ-01	BIOPAC MP 150	Equivital EQ-01
Sampling Freq.	200 Hz	256Hz	1KHz	256Hz
Exp. Conditions	Rest	Office activities	Picture Inspection	Video Game
Principle	Two recording sessions of the same individuals (one month apart), in order to examine the permanence of the biometric features.	Long ECG recordings during regular every day activities, in order to examine instantaneous variations of the signal.	Induction of passive arousal using the International Affective Pictures System [72].	Induction of active arousal using a shooting video game of increasing level of difficulty.

Table 3.1: Summary of ECG Databases

Chapter 4

HeartID: Method and Application Frameworks

4.1 Problem Statement

This chapter presents the *HeartID* recognition system which is based on the ECG biometric. The core of this system is the Autocorrelation / Linear Discriminant Analysis (AC/LDA) algorithm that has been originally proposed in [9] and later extended to cardiac arrhythmia settings in [8]. Information on the prototype *HeartID* system can be found in [82].

In *HeartID*, the AC is used for feature extraction, as it naturally explores the quasi-periodic property of the ECG signal. The LDA is a machine learning algorithm that performs feature selection in a class-optimal way. The two algorithms work in synergy so that the recognizer "learns" the idiosyncratic properties of every subject's ECG. However, during collection, noise and recording artifacts may be presented in the signal that the filters are not able to remove. In biometric recognition, it is important that the systems are empowered with a quality assessment tool that allows them to reject a poor reading rather than risking a low confidence decision. Given the properties of the AC/LDA

feature extractor, this chapter presents a solution to automatic quality assessment that boosts the overall accuracy of the system.

Furthermore, this chapter discusses the next step to the deployment of the AC/LDA approach in real-life systems. Since machine learning is incorporated to the recognition framework, the overall architecture needs to be designed in a way that ensures viability in real-life settings. Three frameworks are explored, each leading to a distinct application environment: 1) small-scale recognition 2) large-scale recognition and 3) recognition in distributed systems. The AC/LDA algorithm is adapted to each scenario based on the requirements of the application environment. To adapt the algorithm, we first need to answer the following questions:

1. *Is the population of the enrollees known at the time of system training?*
2. *What is the relative size of the gallery set?*
3. *Do the environment and privacy policy allow the maintainance of a central database of biometric templates?*
4. *Does the system allow communication with the server where the recognizer is trained?*

The answers to the above questions have direct impact on the AC/LDA recognition scheme. For instance, if general access control is envisioned (for example for subway access), it is not practically feasible to train the LDA algorithm on the recordings of the particular enrollees. This chapter will address these concerns by introducing the concept of *generic training* for the ECG signal, as well as by providing solutions that can personalize the recognizer to the intrinsic variability of every individual.

4.2 Pattern Recognition for ECG Biometrics

In this section, the original AC/LDA [9] algorithm is briefly presented in order to extend it to a quality enhanced version. The fact that the autocorrelation is used for feature

extraction is central to the proposed quality assessment approach. More specific, since the discriminative power of the ECG lies within its second order statistics, a factor that guarantees sufficient repeatability is presented to ensure signal quality.

4.2.1 The AC/LDA Algorithm

ECG biometrics is essentially a pattern recognition problem, comprised of three distinct steps i.e., *pre-processing*, *feature extraction* and *classification*.

Pre-processing. The ECG data in raw format contain both high (powerline interference) and low frequency noise (baseline wander) that needs to be eliminated. Baseline wander is caused by low frequency components that force the signal to extend away from the isoelectric line. The source of this kind of artifacts is respiration, body movement or inadequate electrode attachment. Powerline interference is generated by poor grounding or conflicts with nearby devices.

To reduce the effects of noise, a butterworth bandpass filter of order 4 is used. Based on the spectral properties of each wave in the heart beat, the cutoff frequencies of the filter are 1Hz and 40Hz. The order of the filter and the pass-band frequencies are selected based on empirical results [9, 83].

Feature Extraction. As mentioned before, the core of the proposed feature extraction method is the autocorrelation (AC) of ECG signals. The rationale for AC is that it captures the repetitive property of the ECG signal in a way that only significant, iterative components contribute to the waveform i.e., the P wave, the QRS complex and T wave. By analyzing the AC, incidental patterns of low discriminative power are attenuated, while persistent components of discriminative power are brought to light.

The syllogism behind AC with respect to fiducial points detection, is that it blends into a sequence of sums of products, ECG samples that would otherwise need to be sub-

jected to fiducial detection. Furthermore, the AC allows a shift invariant representation of similarity features over multiple cycles. The AC can be computed as:

$$\widehat{R}_{xx}[m] = \sum_{i=0}^{N-|m|-1} x[i]x[i+m] \tag{4.1}$$

where $x[i]$ is the windowed ECG for $i = 0, 1...(N - |m| - 1)$, and $x[i + m]$ is the time shifted version of the windowed ECG with a time lag of $m = 0, 1, ...(M - 1); M << N$. Even though the major contributors to the AC are the three characteristic waves, normalization is required because large variations in amplitudes appear, even among the windows of the same subject. In addition, only a segment of the AC vector propagates to LDA, as defined between the zero lag instance and up to approximately the length of the QRS complex. This is because this complex is the least affected by heart rate variability [17].

An AC vector can be used directly for classification. However, it is important to further reduce the intra-subject variability in order to control false rejection. In addition, depending on the sampling frequency of the ECG signal, the dimensionality of an AC window can be considerably high. For this reasons the Linear Discriminant Analysis (LDA) is recruited for dimensionality reduction.

The LDA is a well-known machine learning method for feature extraction. Supervised learning is performed in a transform domain so that the AC vector's dimensionality is reduced and the classes are better separable. The remaining discussion is based on the following definitions:

- Let U be the number of classes i.e., the number of subjects registered in the system.
- Let U_i be the number of AC windows for a subject (class) i, where $i = 1...U$.
- We define as $\mathbf{z}_{ij}$ an AC window j, where $i = 1...U_i$ and $j = 1...U$.
- Let $\mathcal{Z}_i$ be the set of AC windows for a subject (class) i, defined as $\mathcal{Z}_i = \{\mathbf{z}_{ij}\}_{j=1}^{U_i}$.

- Let $\mathcal{Z}$ be a training set consisting of all AC windows of all subjects i.e., $\mathcal{Z} = \{\mathcal{Z}_i\}_{j=1}^{U}$.

Then a set of K feature basis vectors $\{\psi_m\}_{m=1}^{K}$ can be estimated by maximizing the Fisher's ratio which is equivalent to solving the following eigenvalue problem:

$$\arg\max_{\psi} \frac{|\psi^T \mathbf{S}_b \psi|}{|\psi^T \mathbf{S}_w \psi|} \tag{4.2}$$

where $\psi = [\psi_{\mathbf{1}}, ..., \psi_{\mathbf{K}}]$, and $\mathbf{S_b}$ and $\mathbf{S_w}$ are the inter-class and intra-class scatter matrices respectively, computed as follows:

$$\mathbf{S}_b = \frac{1}{N} \sum_{i=1}^{U} (\overline{\mathbf{z}}_i - \overline{\mathbf{z}})(\overline{\mathbf{z}}_i - \overline{\mathbf{z}})^T \tag{4.3}$$

$$\mathbf{S}_w = \frac{1}{N} \sum_{i=1}^{U} \sum_{j=1}^{U_i} (\mathbf{z}_{ij} - \overline{\mathbf{z}}_i)(\mathbf{z}_{ij} - \overline{\mathbf{z}}_i)^T \tag{4.4}$$

where $\overline{\mathbf{z}}_i = \frac{1}{U_i} \sum_{j=1}^{U_i} \mathbf{z}_{ij}$ is the mean of class $\mathcal{Z}_i$ and N is the total number of windows and $N = \sum_{i=1}^{U} U_i$.

The maximization of Fisher's ratio is equivalent to forcing large separation between projected ECG windows of different subjects, and small variance between windows of the same subject. The LDA finds ψ as the K most significant eigenvectors of $(\mathbf{S}_w)^{-1}\mathbf{S}_b$ which correspond to the first K largest eigenvalues. A test input window $\mathbf{z}$ is subjected to the linear projection $\mathbf{y} = \psi^T \mathbf{z}$, prior to classification.

It is important to note that ECG biometrics benefit from supervised machine learning approaches more than other biometric modalities. This is because of this signal's dynamic and time-dependent nature which leads the biometric to exhibit higher intra-class variability than traditional biometric modalities. With the LDA one can essentially control false acceptance and false rejection.

Classification. For biometric matching, input and gallery templates are associated using the Euclidean distance as a measure of dissimilarity, while the final decision is made

upon voting of k-Nearest Neighbors. The normalized Euclidean distance is computed as follows:

$$D(\mathbf{y_1}, \mathbf{y_2}) = \frac{1}{K}\sqrt{(\mathbf{y_1} - \mathbf{y_2})^T(\mathbf{y_1} - \mathbf{y_2})} \tag{4.5}$$

where K is the dimensionality of the feature vectors. For a U class problem, LDA can reduce the dimensionality of the feature space to U-1 due to the fact that the rank of the between class scatter matrix cannot go beyond U-1. Factor V is there to assure fair comparisons for different dimensions that $\mathbf{z}$ might have.

4.2.2 Quality Assessment with the Periodicity Transform

While the main ingredient of proposed biometric recognition system is the AC/LDA algorithm, this thesis presents an extension of the basic approach with the incorporation of a signal quality assessment methodology. The enhanced AC/LDA feature extraction algorithm is two-fold. First, the system utilizes the periodicity transform (PT) to judge the matching validity of the signal, and second, a localized matching approach is adopted to address further the heart rate variability issues.

Following the above description of the basic AC/LDA algorithm, the analysis is herein presented in three steps.

Pre-processing. After application of the previously discussed Butterworth filter, the enhanced AC/LDA algorithm estimates a quality measure, Q_i, for an input ECG segment i. Q_i expresses the confidence of the system in analyzing a valid signal i.e., an ECG segment free of major artifacts.

To this end, the periodicity transform (PT), proposed by Sethares *et al.* [84], is used to project the signals into a sum of periodic sequences. PT is a data driven methodology that can identify the *best* set of bases, each of which describes an inherent periodicity of the ECG. The rationale for PT in this work is that the ECG is a quasi-periodic signal with a specific repetitive pattern, and any variation from this, is usually attributed to

recording errors or noise effects. The stronger the periodic component of the waveform the more likely it is for the signal to be free of artifacts. Furthermore, the AC/LDA method relies on ECG's repetitive nature for feature extraction. It is because of quasi-periodicity that the AC can capture in detail information about the waves that form a heart beat. However, a corrupted signal will not conform with its class general behavior. PT can essentially identify these signals and assign them with a low quality value.

For the PT a set of periodic and orthogonal basis functions is defined as follows:

$$\delta_p^s(j) = \begin{cases} 1, if(j-s) \bmod p=0 \\ 0, otherwise \end{cases} \tag{4.6}$$

where p is the period and s is a time shift, with $s = 0, 1, ...p-1$. According to PT, P_p defines a set of (time-shifted) periodic sequences of period p. The projection of an ECG signal x, on to the periodic subspace P_p can be expressed as a linear combination of the periodic elements $\delta_p^s(j)$ as follows:

$$\alpha_s = \frac{1}{N} \sum_{n=0}^{N-1} x(s+np) \tag{4.7}$$

$$\pi(x, P_p) = \sum_{n=0}^{N-1} \alpha_s \delta_p^s \tag{4.8}$$

where $\pi(x, P_p)$ is the projection of x on subspace P_p. In essence, the PT seeks for the best periodic characterization of the signal in-hand. Algorithmically, the search for the best period is iterative, with faster periods being detected first:

1. A signal x is projected on the periodic subspace P_p, to acquire $\pi(x, P_p)$

2. If the periodicity is significant, it is removed from the data, $r_p = x - \pi(x, P_p)$.

3. r_p is searched for larger periodicities.

The simpler way to perform these operations is to seek small periodicities p first, and to iteratively remove them from the signal. Essentially, a period p can be as small as 2

however, it is not particularly meaningful to look for periodicities that are bigger than $N/2$, where N is the length of the signal. The *Small-to-Large* algorithm operates as follows [84]:

`pick threshold` $T \in (0, 1)$

`let` $r_0 = x$

for $p = 2$ to $N/2$ **do**

  $x_p = \pi(x, P_p)$

  if $\frac{\|r - x_p\|}{\|x\|} < T$ **then**

    $r = r - x_p$

  end if

end for

For the AC/LDA enhancement, the norm (power) is first used to select the M most prominent periodicities of the signal, along with the corresponding projections. At a second step, each of the projections is revisited to examine possible decompositions into stronger sub-periodicities that can replace the first. The best period is identified by the highest returned power. For the ECG case, it is expected that this period will correspond approximately to the heart rate. The power criterion utilized during the decomposition may serve as a quality metric. Q_i is therefore defined as:

$$Q_i = \frac{\|x_i - \pi_i(x, P_p)\|}{\|x_i\|} \tag{4.9}$$

where x_i is the ith ECG input and $\pi_i(x, P_p)$ is its periodic projection onto the bases of the best period. Essentially, Q_i describes the comparative energy between the original and the projected onto the best periodicity signal. *Accordingly, Q_i is a measure of how strong the repetition of ECG is, at a period approximating the heart rate.* The higher the Q_i for an input i the greater the confidence about the collected signal. The following discussion explains the way that this metric can be incorporated with the recognition

methodology.

Feature Extraction. In the enhanced AC/LDA algorithm, the AC is divided into n equal segments, and LDA is applied on each of them independently. This allows for finer localization of the information and unequal treatment of the AC segments during matching.

The rationale for unequality is because segments which exhibit higher discrimination and better intra-class stability need to be favored. The closer to the main AC peak, the greatest the contribution of the *QRS* complex, which has the highest voltage amplitude within a heart beat. Furthermore, moving away from the AC's main peak, the effect of this wave diminishes. The *QRS* complex exhibits the least variability under varying heart rates [85]. In other words, the intra-subject variability increases while moving away of the main AC peak. Emphasizing the contribution of the *QRS* complex by weighting segments closer to the AC peak higher, can potentially handle the heart rate variability and boost the performance.

Classification. For identification, every input is processed and matched against all previously designed templates which compose the gallery set. The similarity measure considered, is the Euclidean distance and classification is carried out with the nearest neighbor. In the enhanced AC/LDA algorithm, the LDA coefficients of each of the AC segment are matched against the respective gallery segments. A distance d_n is estimated for every pair, and multiplied with a weight w_n corresponding to the n-th segment. Lower order segments (closer to the AC peak), are weighted higher. All products are averaged to estimate a distance d between a pair:

$$d = \sum_{j=1}^{n} \frac{d_j w_j}{n} \tag{4.10}$$

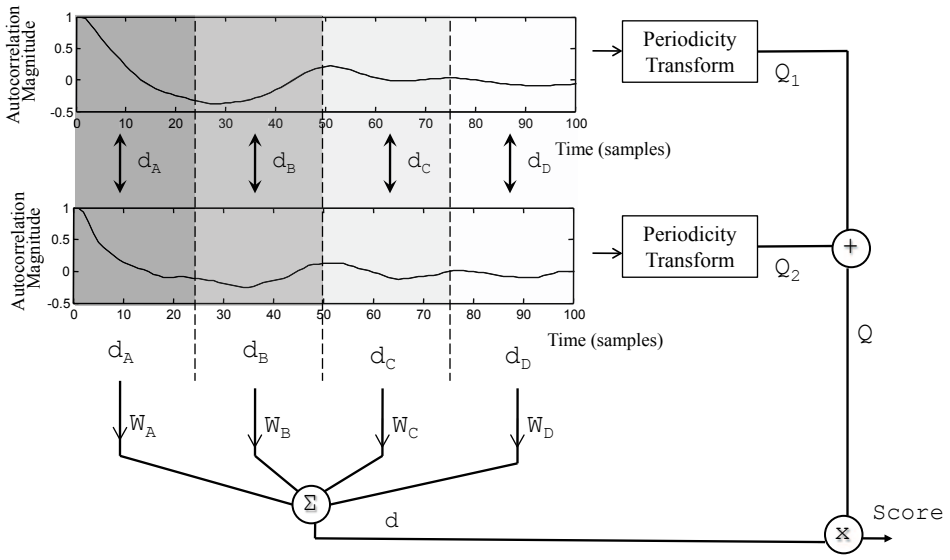


Figure 4.1: Flow diagram of the proposed method. Every input is assigned with a quality measure that contributes to matching. The AC is divided into a number of sections each favored with a predefined weight. The weights W_{A-D} cab be chosen to decrease linearly.

Before finalizing a decision, the pair's distance d is weighted with the inverse mutual quality Q, which is defined based on the respective quality measures Q_i, as follows:

$$Q = \frac{1}{\frac{1}{m}\sum_{i=1}^{m} Q_i} \tag{4.11}$$

where $m = 2$, when estimating the quality of a pair. The inverse quality is used in order for d and Q to be distance descriptors of the final score:

$$Score = d \times Q \tag{4.12}$$

A threshold on the score allows the system to accept or reject an identity claim made by a user. This algorithm is summarized in steps in Table 4.1 and a flow diagram is provided in Figure 4.1.

Testing Algorithm

Step 1: Perform PT on input x_i and estimate Q_i.

Step 2: $\widehat{R}_{xx}(m) = \sum_{k=0}^{N-|m|-1} x_i(k)x_i(k+m)$.

N is the length of x_i and m is the AC lag.

Step 3: $\widehat{R}_{xx} = \frac{\widehat{R}_{xx}}{\widehat{R}_0}$ (Normalization)

Step 4: $r_{xx}(t) = R_{xx}(m)$ $0 \leq m \leq k$, where k is the length of a heart beat.

Step 5: Divide $r_{xx}(t)$ in n equal segments of length t': $r^1_{xx}(t')...r^n_{xx}(t')$.

Step 6: Project(LDA) $r^1_{xx}(t')...r^n_{xx}(t')$ in a lower dimensional space: $s^1(t'')...s^n(t'')$ where $t'' < t'$

Step 7: Estimate the distance from the gallery features $g^1(t'')...g^n(t'')$, $d_1 = d(s^1, g^1)...d_n = d(s^n, g^n)$ where $d(.)$ is the Euclidean dist.

Step 8: Weight d_j with w_j.

Step 9: $d = \sum_{j=1}^{n} \frac{d_j w_j}{n}$

Step 10: $Score = d/mean(Q_i, Q_g)$ where Q_g is the respective quality factor of the gallery subject.

Table 4.1: Basic steps of the proposed framework.

4.3 Application Frameworks

A typical setup for biometric recognition involves an enrollment and a recognition stage. Under both modes of operation, ECG is filtered and subjected to feature extraction. However, the exact manner in which this is done depends highly on the requirements of the envisioned application. Central to ECG biometrics is the establishment of the learning algorithm (i.e., the LDA) on bases that can accommodate practical recognition scenarios.

This section presents three unique implementation frameworks of the ECG biometric recognition technology. For simplicity, the training stage of the LDA is separated from enrollment to accommodate the description of open recognition environments i.e., settings where the gallery set is not available at the time of system training.

4.3.1 Scenario A. Small-scale Recognition Environments

In small scale applications, where the individuals to be recognized are known a priori, the variability of the signal among the monitored population is learned by training the learning algorithms on recordings from the particular enrollees. A block diagram of this system is depicted in Figure 4.2.

During enrollment (which in this case plays the role of training as well), an ECG reading is acquired from every individual, filtered in order to remove noise, and subjected to AC estimation. The autocorrelated signals of all enrollees are saved in a central database. This set of signals is then used as input to drive the learning algorithm (i.e., the LDA). This operation will statistically minimize the intra-subject variability and maximize the inter-subject one. The transformation matrix (ψ) provided by this step guarantees that when reapplied on an ECG reading (during recognition) of one of the enrollees, it will transform the signal in a way that data from different users are optimally separated.

This scenario may operate under either the identification or verification mode of operation. For the identification mode, a new reading is classified as one of the pre-enrolled identities. Accordingly, for the verification mode, an identity claim is first made, which is used to retrieve the respective recording from the database. This recording is then matched against the input, in order for the system to either accept or reject the claim.

The key features of Scenario A is the known population of enrollees and the possibility of maintaining a central server with the biometric templates. Examples of use cases for

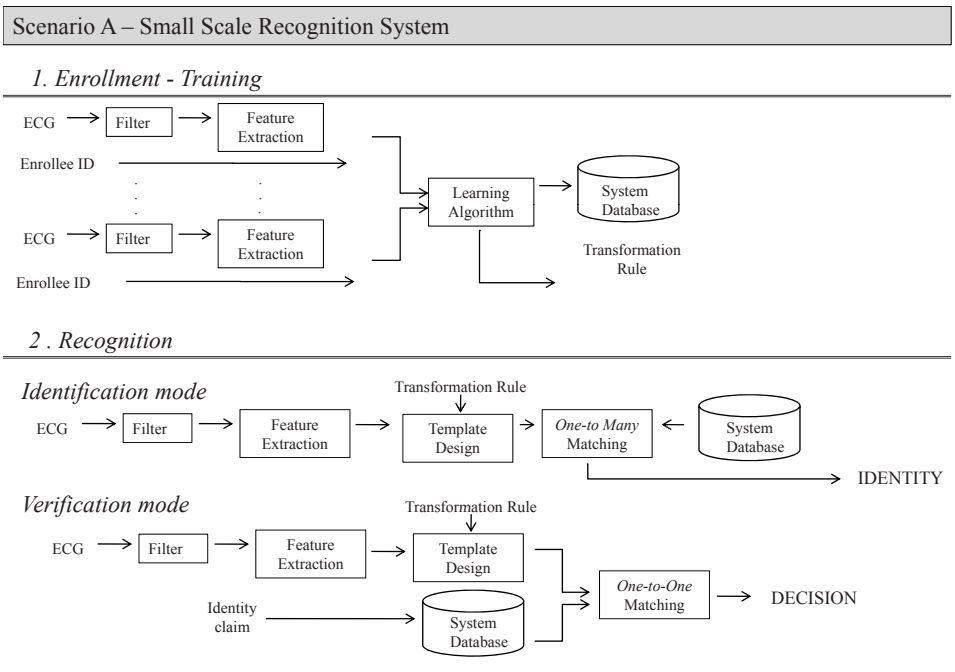


Figure 4.2: The two stages of an ECG based recognition system in closed environments i.e., cases where the pool of enrollees is known prior to LDA training.

this systems include:

1. *Access control within the facilities of a company.* The company invites the employees to a signal collection session, and an ECG sample is acquired from every employee. The LDA is trained *offline* to learn the ECG morphologies of the particular employees. The system is then used for physical access control with the following possibilities:

   - Identification: When an employee requests access to particular facility, a new sample of the biometric is collected, and matched against the database that is stored centrally. The answer of the recognizer can be one of the three:

     (a) The identity information of the employee (answering the question *Who is this employee?*).

     (b) The clearance level of the particular employee.

(c) A Yes/No response equivalent to a watch-list search i.e., answering the question *Is this person an employee?*

- Verification: An employee requests access to a facility and at the same time presents credentials that make an identity claim (for example a name badge, or ID card). The biometric sample is collected and compared against the biometric template that corresponds to the claimed identity. The system replies with a YES / NO answer.

2. *Field agent authentication.* In welfare monitoring environments, for instance for soldiers whose vital signals are being monitored continuously from a central authority, ECG biometrics may validate their identities continuously to avoid agent impersonation. Enrollment is performed once and then recognition may take the form of either identification or verification:

   - Identification: The monitoring authority receives vital signals from unknown sources. The incoming medical signal is matched against a number of pre-enrolled biometric templates and the identity of the transmitting agent is established.

   - Verification: When extra identifying credentials can be employed, for example an ID number of the sensor unit, the respective ECG template is used for biometric matching at the receiver and a YES/ NO answer is provided. A system in this mode answers the question *Is the monitored agent the person I expect him/her to be?*

4.3.2 Scenario B. Large-scale Recognition Environments

This framework addresses the problem of ECG based recognition in large-scale recognition settings. The challenge with such systems is that the morphologies of the enrollees' ECG signals are not known at the time of LDA training. In addition, in the majority

of such environments the interest is in protecting a particular user's template non only from attacks initiated by users (known to the system), but also from third parties that are not registered with the system.

From a machine learning point of view, the objective in large-scale systems in to *optimally reduce the intra-class variability while learning patterns of the general population.* The general population is herein referred to as the *generic pool* and it is defined as an anonymous collection of ECG samples from a large number of individuals. The purpose of the generic pool is to create a paradigm of ECG morphologies that an intruder might be presented as.

The large scale recognition framework is depicted in Figure 4.3. Training is the stage where the generic pool is formed and the LDA algorithm is trained. At this step, a transformation matrix (ψ) is produced to project an arbitrary ECG input in a space of high inter-subject variability. Subsequently, during enrollment, ψ is used for projection of the input ECG features, which results in a biometric template design (y). Given the prior training on an anonymous pool, the enrolled template is statistically protected against a variety of other ECG morphologies. Similar to Scenario A, recognition takes the form of either identification or verification, depending on the envisioned application environment.

The key features of Scenario B is the unknown population of enrollees at the time of LDA training, and that recognition is performed centrally, on a server where the biometric templates are stored.

An example environment for the application of this framework is identification for subway access. The generic pool is created *offline* and the system is trained in order to acquire a transformation matrix (ψ). ψ is then used by the subway authority to design biometric templates. When a user pays a fare, a sample of the ECG signal is collected, and a template is designed for that individual. The template is stored centrally on a server and recognition is then performed with the following possibilities:

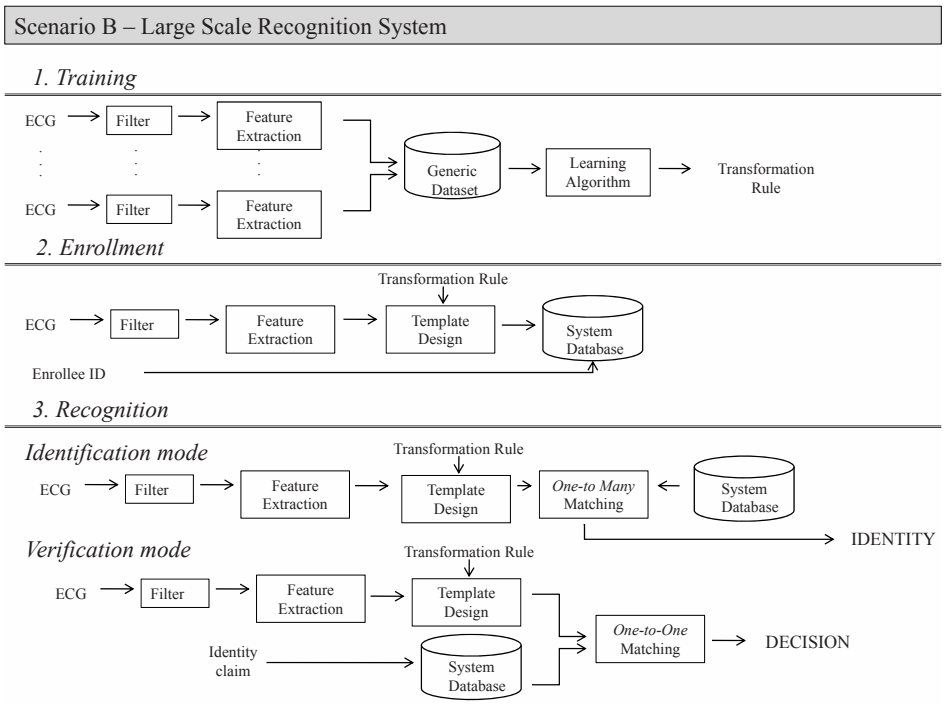


Figure 4.3: The three distinct stages of general access control.

- Identification: This takes the form of *watch-list* based operations whereby a new biometric sample is collected, compared against the pre-enrolled ones and the output is a YES/NO decision based on one-to-many matches. Essentially, this system answers the question *Has this person paid the legitimate fare?*

- Verification: During recognition the user presents a subway pass which is linked to a biometric template that is stored centrally. A new ECG signal is collected and matched against the template that the card indicates. The answer of the system is YES/NO. This system answers the question *Is this user the legitimate card holder?*

4.3.3 Scenario C. Security in Distributed Systems

This framework can be used in either small or large scale recognition environments i.e., within settings of known or unknown population of enrollees. The idea behind the distributed system is that recognition takes place locally on a device such as a smart-card

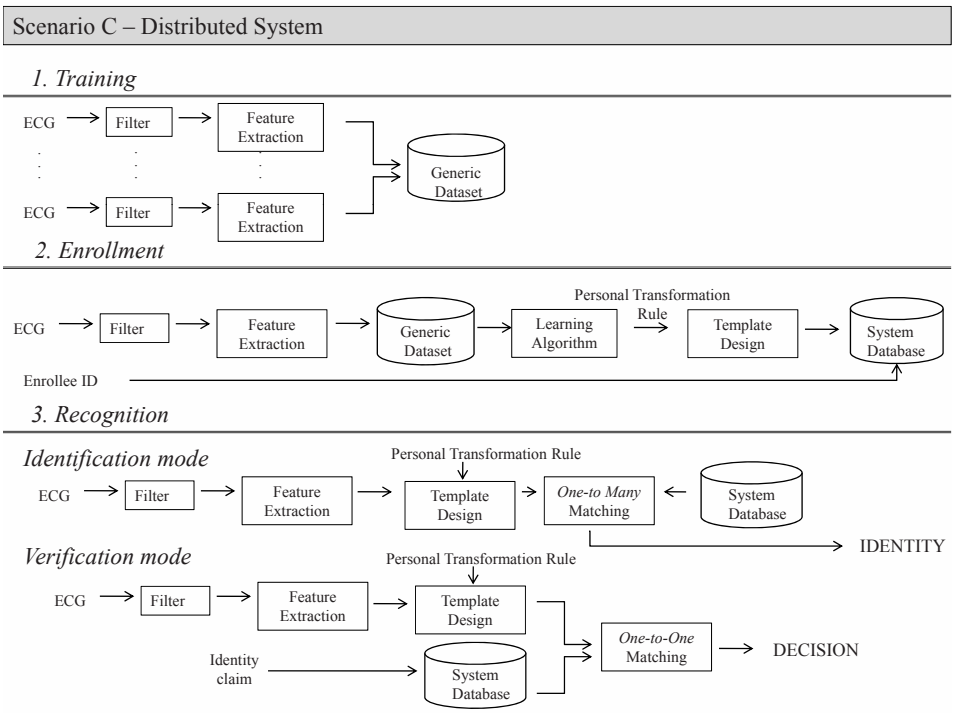


Figure 4.4: The three distinct stages of ECG-based recognition in distributed systems.

or smart-phone. This setup has the benefit of allowing the recognizer to be personalized to every user. A block diagram of this system is depicted in Figure 4.4.

The training phase is separated from enrollment because in this Scenario the transformation matrix (ψ) is personalized for every user. Since in most distributed recognition environments, the outliers may be subjects that are unknown to the system (for instance an illegitimate user that tries to be authenticated on a smart-phone), the generic pool is once more employed as a way to handle false acceptance.

At an algorithmic level, the personalization at the enrollment stage takes the form of learning one's ECG sample against the generic pool. More specific, given a generic dataset of anonymous ECG recordings, the AC is first estimated for each of them independently. This results in a number of AC segments, $\Phi(m)$, against which an input AC feature vector, $\phi_{input}(m)$, is to be learned. $\phi_{input}(m)$ is the autocorrelated ECG sample of the legitimate user at the time of the enrollment on the smart device.

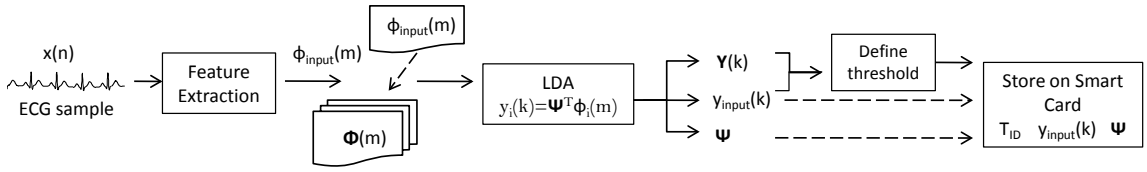


Figure 4.5: The enrollment pipeline for the distributed verification framework.

Let the number of subjects (classes) in the generic pool be C. The training set will then involve $C+1$ classes as follows:

$$\Phi(m) = [\Phi_1(m), \Phi_2(m)...\Phi_C(m), \Phi_{input}(m)] \tag{4.13}$$

and for every subject i in $C+1$, a number of C_i AC vectors are available:

$$\Phi_i(m) = \{\phi_v(m)\}_{v=1}^{C_i} \tag{4.14}$$

Although multiple recordings per subject are optional, the discriminant will perform better when trained on more than two instances of the biometric per subject. Since this is only required at the enrollment stage, it does not affect the overall waiting time of the recognizer.

Given $\Phi(m)$, LDA will find a set of k feature basis vectors $\{\psi_v\}_{v=1}^{k}$ by maximizing the ratio of inter-class and intra-class scatter matrix. As mentioned earlier in Section 4.2.1, this maximization is equivalent to solving the following eigenvalue problem:

$$\arg\max_{\psi} \frac{|\Psi^T \mathbf{S}_b \Psi|}{|\Psi^T \mathbf{S}_w \Psi|}, \Psi = \{\psi_1, ..., \psi_k\} \tag{4.15}$$

where $\mathbf{S}_b$ and $\mathbf{S}_w$ are inter-class and intra-class scatter matrices. Given the personalized transformation matrix Ψ, a feature vector is projected using:

$$Y_i(k) = \Psi^T \Phi_i(m) \tag{4.16}$$

where eventually $k << m$ and at most C.

An advantage of distributed recognition, that operates under the verification mode of operation, is that every smart device can be optimized experimentally for the intra-class

variability of the particular user. On a typical ROC plot of False acceptance and False Rejection rates (FAR and FRR), the latter depends only on the intra-class variability of the feature vectors. On the other hand, FAR is a measure of inter-class variability of the feature space.

Therefore, by choosing the smallest distance threshold at which an individual is authenticated, also guarantees minimum FA. Essentially, rather than imposing universal distance thresholds to all users, we propose that every device is "tuned" with a *personalzied threshold* T_{ID} based on the variability of the ECG recordings at the time of the enrollment. This can be done with cross fold validation of the distances among the various enrolled templates. Finally, in every device the following verification triplet is saved $\{\Psi, Y_{input}[k], T_{ID}\}$. This procedure is graphically depicted in Figure 4.5.

The key features of Scenario C are that it can be applied for recognition in both known or unknown population settings, as well as that verification can be distributed i.e., performed locally on a smart device where the biometric template is also saved. An example use case for this system is as follows:

Identity Verification on a cell phone. Assuming limited processing and storage capabilities of a cell phone, the generic pool of ECGs is stored centrally on a server with which the device needs to communicate. Following the steps of Figure 4.5, the device collects an ECG sample and computes the autocorrelation (feature extraction). The feature vector is transmitted to the server where the respective features of the generic pool, $\Phi(m)$, are saved. The new vector, $\Phi_{input}(m)$ is appended to $\Phi(m)$, and the LDA is trained upon it. The resulting transformation matrix, Ψ, describes a subspace where the ECG of the particular user is discriminated from various morphologies which may take the form of an intruder. In other words, $\Phi(m)$ is a rule meant to protect the particular user against the general population. A user specific threshold, the transformed feature space, $y_{input}(k)$ (k ¡¡ m), and Ψ are transmitted back to the device, where they are saved.

During recognition, the device can verify the user by using a newly collected ECG

signal and the personalized matrix Ψ. The processed signal is matched against the stored biometric template and a YES/NO answer is acquired.

4.4 Performance Evaluation

This section reports the performance of the proposed biometric recognition system. The objective of the following experimentation is to demonstrate the accuracy of the enhanced AC/LDA system as applied in small-scale, large-scale and distributed recognition systems. In addition, this section provides recognition results that simulate ECG biometric authentication in monitoring environments. All frameworks of this section were tested using the Short and Long-Term ECG Databases (Chapter 3.5).

4.4.1 Quality Assessment Results

The following evaluation was carried out for signals in the short-term recording database (52 subjects) under a small-scale recognition framework. This framework (Scenario A) is the most typically used experimental setup in biometric recognition, as the machine learning algorithm is trained on the enrollees' ECGs which are assumed to be available.

Every 3 min ECG was divided into a number of 5 sec windows. The 5 seconds were chosen as a feasible compromise between the system's waiting time and meaningful feature extraction. Overall, 1172 ECG windows, from 52 subjects, were used for testing. The normalized AC of each window was estimated and divided into 4 segments for matching, according to the requirements of the enhanced AC/LDA algorithm. The reader should note that recordings of higher sampling rates might require more segments, to allow for better localization.

The periodicity transform was computed for every ECG window in order to estimate the quality measure Q_i. Figure 4.6 demonstrates its power in detecting outliers. After screening with the PT based quality measure, the AC segments conform approximately

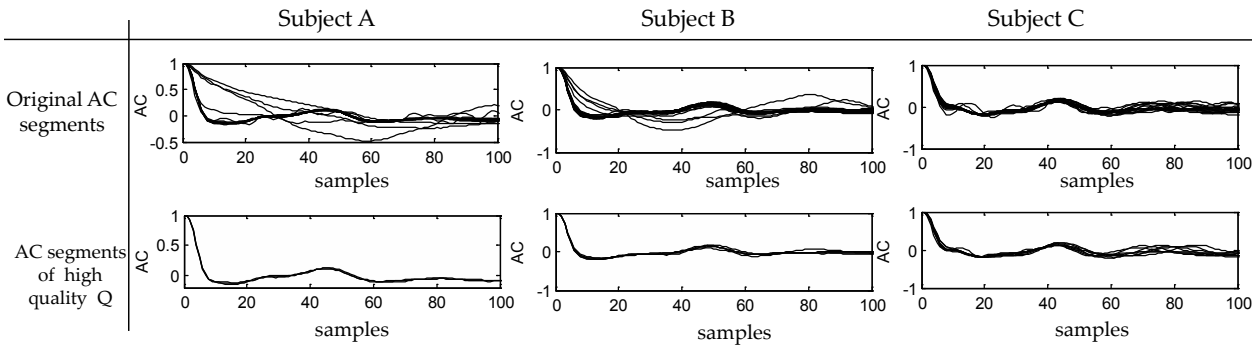


Figure 4.6: Performance of the PT quality factor. AC segments screened with a quality threshold (0.8).

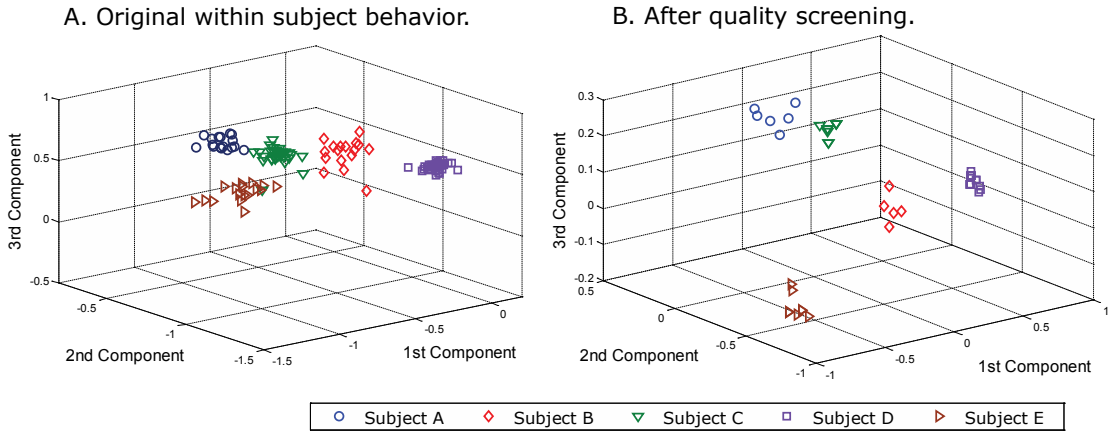


Figure 4.7: Three principal components of AC for 5 different subjects A) before and B) after quality screening.

to the same distribution for each subject in the gallery set. This can effectively decrease the within subject variability, as depicted in Figure 5.4, where three principal components (estimated with PCA for demonstration purposes) of the AC are plotted for a number of subjects. In comparing with the baseline AC/LDA system, the within-class distance drops by *25%* (for a quality threshold of 0.8), when using the PT quality extension.

The proposed system was first tested under the verification mode of operation. The objective is to answer the question: *Is the subject who he/she claims to be?*, thus the user makes a positive claim to an identity. This kind of recognizers may commonly do one of

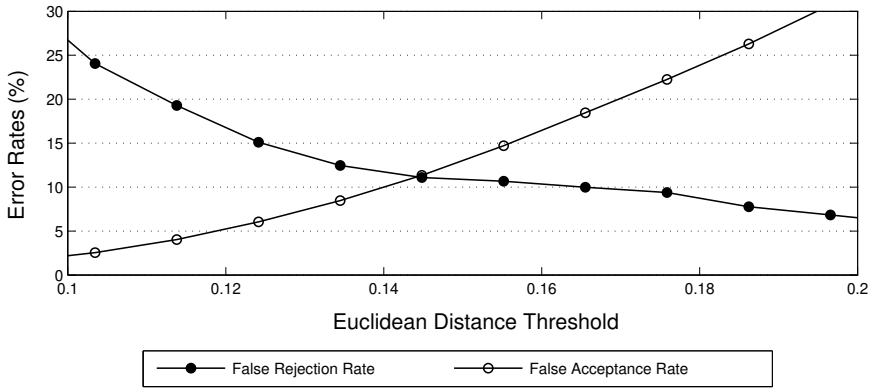


Figure 4.8: Tradeoff between false acceptance and rejection rates for various decision thresholds.

the following mistakes:

1. Mistakenly accept an identity claim made by an intruder (false acceptance (FA))

2. Mistakenly reject an identity claim made by a legitimate user (false rejection (FR)).

Since the final decision is based on a distance threshold between the two templates, there is a tradeoff between these two error rates. Figure 4.8 illustrates the performance of the system in terms of FA and FR rates. One should note, that if the quality of the readings is ignored, and the system is tested without the PT extension, the EER (i.e., the false rate when FA and FR are equal) may increase by 2% for the current dataset.

Alternatively, when the system is tested under the identification mode of operation, thus answering the question: *Who is this subject?*, the true positive performance, is *81.48%* for 5 sec ECG and can reach *92.3%* for 3 min recordings (using majority voting of the respective 5 sec windows). This performance is comparable with other works in the field [37, 34, 48] but for more subjects (52), longer recordings, and without the additional complexity of fiducial points detection, which naturally leads to rejection cases.

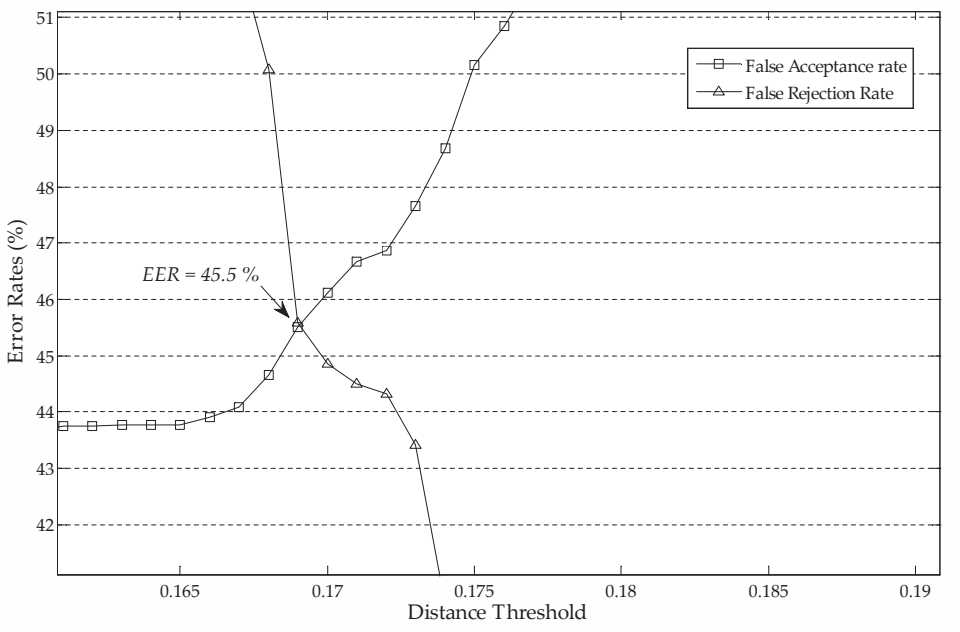


Figure 4.9: ROC plot depicting the performance of Scenario B. The EER is *45%*.

4.4.2 Training on the Generic Pool

This section presents the performance of Scenario B for large scale recognition environments. In the following, the evaluation was performed on the short-term recording database. The generic pool was formed using signals from the 36 volunteers who participated to the experiment only once. Every signal was partitioned into segments of 5 sec length each, resulting in a total of 1296 samples. For the 16 volunteers that two recordings were available, the earliest ones were used for enrollment (gallery set with a total of 576 samples) and the latter for testing (same number of samples).

The ECGs of the generic pool were subjected to autocorrelation estimation and then to discriminant analysis with the LDA. The LDA yielded a transformation matrix, which was subsequently used to transform the ECG signals of the gallery and testing sets. All enrollees were processed with the same generic transformation matrix. Matching was then carried out using the Euclidean distance as the similarity measure.

A threshold on the distance between the two matches allows the system to either

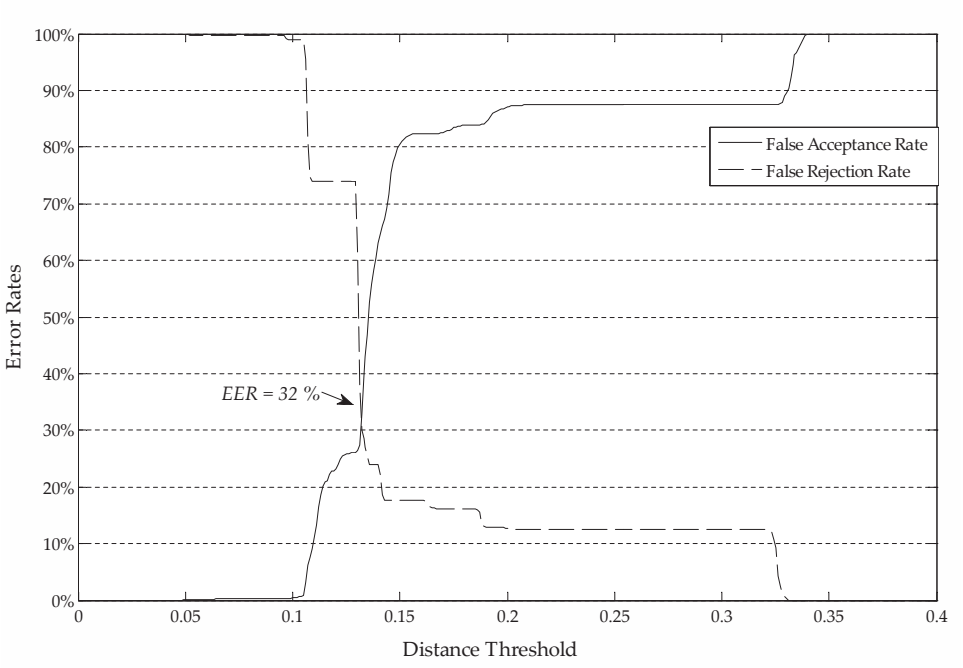


Figure 4.10: ROC plot when imposing universal recognition thresholds in Scenario C.

accept or reject the pair. Figure 4.9, demostrates the tradeoff between false acceptance and rejection for various thresholds. The EER reaches dramatically *45%*. This is generally an unacceptable performance for real-life recognition systems, as it suggests that an arbitrary ECG has equal chances of been mistakenly accepted or rejected by the system.

This performance, however, is expected as in this Scenario the LDA is trained on ECG waveforms that are not included in testing. In other words, the system did not "learn" the particular waveforms of the enrollees and thus is not capable of handling false rejection. On the contrary, as one increases the size of the generic set false acceptance is expected to decrease.

4.4.3 Personalized Recognition

This section reports the performance of Scenario C, for ECG biometric recognition in distributed systems with the ability of personalization. The subsequent analysis was

Subject ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
EER	7%	21%	13%	9%	12%	17%	3%	3%	5%	0%	5%	16%	23%	13%	13%	0%
Improvement	38.5%	24.5%	32.5%	36.5%	33.5%	28.5%	42.5%	42.5 %	40.5%	45.5%	40.5%	29.5%	22.5%	32.5%	32.5%	45.5%

Table 4.2: EER in Scenario C for individual subjects in the testing set. The average EER is *10%* and the standard deviation *7.13*

performed on the short-term recording database. Since this scenario also requires a generic pool, signals from the 36 volunteers that participated to the experiment only once were used to form it.

Upon estimation of the AC for all records in the generic pool, each of the enrollees' autocorrelated recordings were appended to pool individually, and a new LDA was trained for every enrollee. The performance each recognizer was then tested with matches against the respective subject's testing recordings. To estimate the false acceptance rate, the remaining enrolles (i.e., subjects who did not participate in the generic pool), acted as intruders to the system. This subset of recordings is unseen to the current LDA, and thus constitutes the unknown population of potential intruders.

Figure 4.10 demonstrates the tradeoffs between false acceptance and rejection when the same threshold values are imposed for all users. The equal error rate is *32%*. This performance, although better than Scenario B, is yet unacceptable for a viable security solution.

When verification is performed locally on a smart device, one can take advantage of the fact that every device can be personalized to a particular individual. This treatment controls the false rejection since the matching threshold is "tuned" to the biometric variability of the individual. Table 4.2, reports the EER performance per subject for the case of personalized thresholds. On average, the EER deceases to *10%* (standard deviation *7.13*), with a significant number of subjects exhibiting EER between *0%- 5%*.

4.4.4 Template Destabilization

This section demonstrates the challenge of ECG biometric recognition due to template destabilization. As mentioned in the introductory sections the ECG signal is affected by both physical and psychological activity. Physical activity increases the heart rate, a problem which is intrinsically solved by using the autocorrelation for feature extraction. On the other hand, the psychological activity on the ECG may be unpredictable.

From a biometric recognition perspective, emotional activity embedded in ECG may adversely affect security. A way to address this issue would be to collect as many ECG readings from a subject as possible during enrollment, in order to cover a range of emotional expressions. However, there is no guarantee that adequate instances were acquired, nor can one reassure that distinct psychological states were experienced during the recordings.

To demonstrate the perils of ignoring time dependency the destabilization of the biometric template was examined with time. The following discussion concerns all three Scenarios, as it essentially shows that *in the absence* of physical activity the ECG waveform may still change.

The long-term recording database (Chapter 3.5) was used in the following experimentation because it offers a large numbers of ECG windows per subject. This database also simulates real life settings of cognitive activity, as the volunteers were monitored during their daily activities.

Every 2-hour recording in the database was segmented into non overlapping windows, each of 5-second length. The correlation between a training subset and the subsequent windows was examined based on the correlation coefficient between the respective autocorrelations. The correlation coefficient value between a reading at time t_r (reference) and any following time instance t_t can be estimated as using:

$$P_{R(t_r)R(t_t)} = \frac{E[(R(t_r) - \mu_{R(t_r)})(R(t_t) - \mu_{R(t_t)})]}{\sigma_{R(t_r)}\sigma_{R(t_t)}} \tag{4.17}$$

where $\mu_{R(t_i)}$ is the mean of the i^{th} ECG reading and $\sigma_{R(t_i)}$ is the respective standard deviation.

Figure 4.11 shows the correlation coefficient values for two different subjects (with five different reference points. The observation that all individual graphs exhibit significant differences clearly illustrates the time-varying nature of ECG, with respect to the second-order statistics. In other words, depending on the starting point in time, the resulting behavior of subsequent windows is characterized with a varying rate of change. Figure 4.12 shows the maximum and minimum correlation coefficient found within the 2 hour recording for all 10 subjects. From this Figure it is clear that the template similarity may drop significantly. The following results will demonstrate how this can affect the recognition accuracy.

For the current experimental setup, the system was tested under a verification mode of operation, but the conclusions can be easily expanded to identification scenarios as well. Every testing ECG input was matched against a claimed identity (the legitimate one when computing authentication rates) using the Euclidean distance as a similarity measure. Using a threshold on this distance, the system decided to either validate or reject the claim.

Figure 4.13 shows the authentication performance averaged over all subjects, for various distance thresholds and time proximities from the training set (the reference). As expected, when time increases, i.e., the time proximity between the training and testing set increases, the authentication rate drops. Thus, the closer to zero the higher the chance for correct verification. Figure 4.13 essentially demonstrates the underlying psychological activity on the ECG signal.

4.5 Chapter Summary

This chapter presented the proposed enhanced AC/LDA algorithm for ECG biometric recognition. In essence, this method provides a factor that can be used to assess the quality of the reading prior to biometric matching. In addition, the enhanced AC/LDA allows for localization of the feature space, as described by the autocorrelation, in a way that features which are more prone to the heart rate variability are less favored.

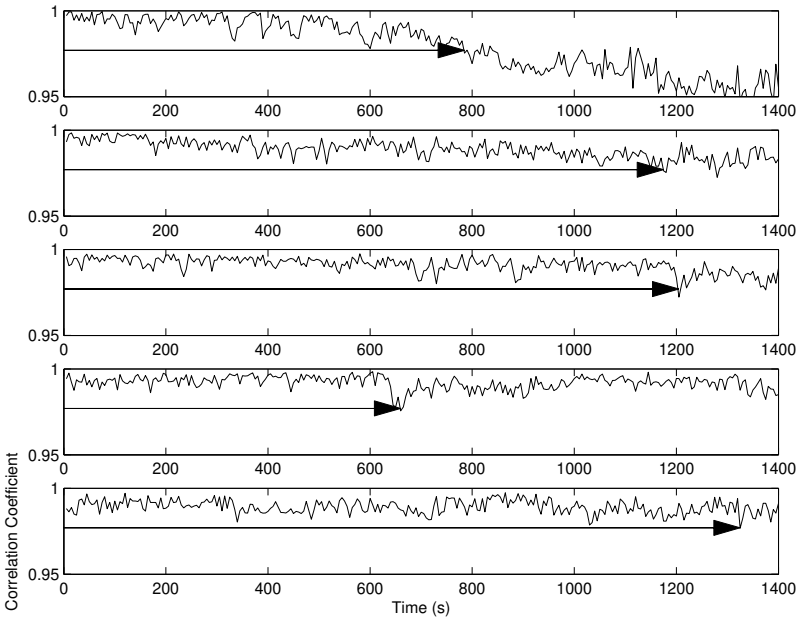
In addition, this chapter presents the three distinct frameworks under which ECG biometric recognition can operate i.e., for small-scale or large-scale recognition environments or in distributed systems. Although all biometric modalities can be used in these settings, the ECG requires special attention in that this time-dependent modality needs machine learning in feature extraction, in order to have a meaningful biometric matching. Therefore, the AC/LDA algorithm has been adjusted to the aforementioned environments.

The more restricted the environment i.e., when the population or enrollees and expected attackers is known, the lower the expected false acceptance and rejection. However, when there is the possibility of performing recognition locally, on a smart device, the recognizer and the authentication thresholds can be personalized with substantial performance benefits.

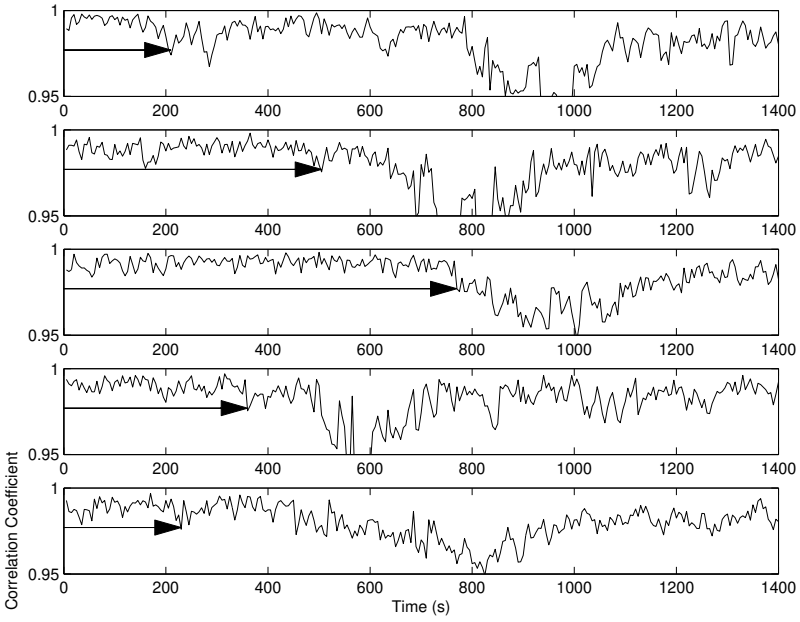
Finally, this chapter demonstrates the perils of ignoring time dependency on the ECG signal. It was observed that with time the discriminative power of the template decreases due to the natural evolution of the signal. This observation is very important for ECG biometrics as it instructs that in the absence of recording noise and physical activity the accuracy of the system may still drop.

The latter conclusion is the basis for the subsequent chapters of this thesis. Although the present simulation results were based on long-term ECG signals, where cognitive activity is assumed, it is not validated yet that template destabilization takes place at the onset of a particular emotional state. Chapter 4 investigates the feasibility of

detecting emotions from ECG signals, with the objective of determining the conditions under which the ECG waveform may change and jeopardize recognition. Subsequently, Chapter 5 brings the attention back to biometric recognition, to provide a solution to this problem using a template updating technique.



(a) Subject 1



(b) Subject 8

Figure 4.11: Correlation coefficient values for two different subjects, with five different reference (starting) points. Corresponding coherence durations (illustrated with arrows) are determined with a threshold or tolerance range of 2.3%, with respect to the starting point.

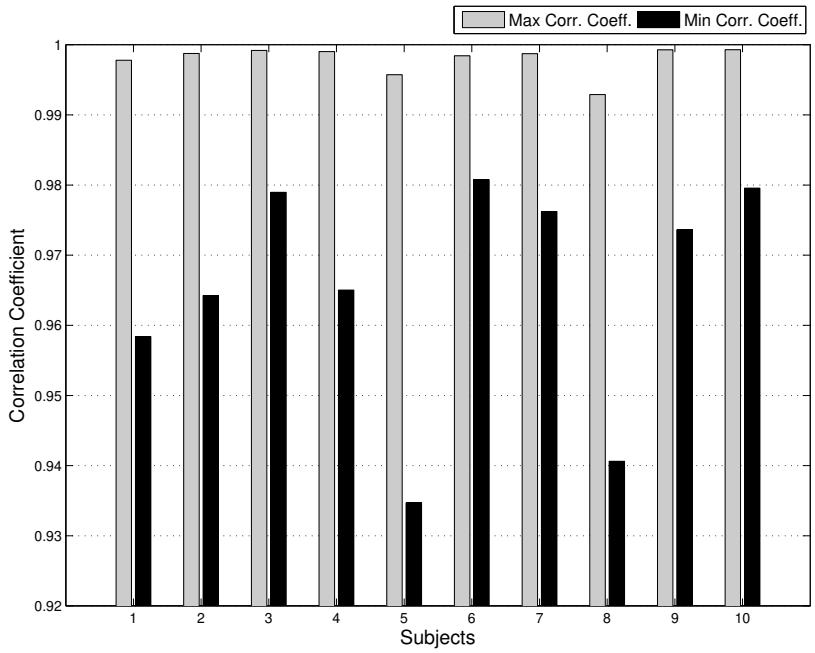


Figure 4.12: Maximum and Minimum correlation found within 2-hour recordings for every subject. Correlation is computed between the first 5 second segment and all the following.

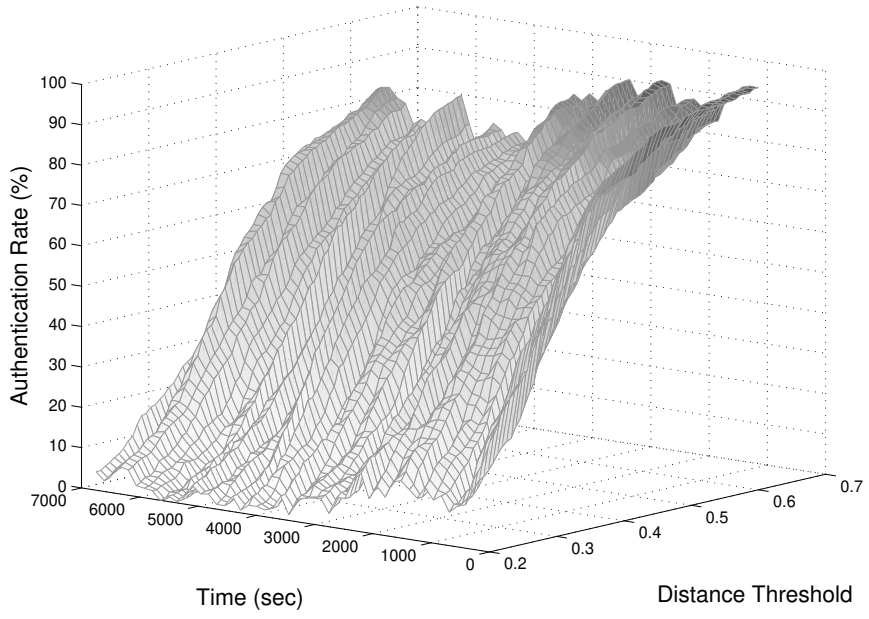


Figure 4.13: Verification performance under template destabilization.

Chapter 5

Affective Patterns of ECG

5.1 Problem Statement

The previous chapter treated the ECG as a biometric signal of physiological origin. The structure of the ECG waves rely heavily of the cardiac properties of every individual and provide the basis for biometric discrimination. However, as discussed in Chapter 3, in the absence of external destabilizing actors (noise or physical activity), the ECG waveform evolves with time. This chapter explores the origin of this evolution.

To this end, the ECG signal is herein investigated from a psychological perspective. The purpose of this work is firstly to, determine the emotional states under which the ECG waveform may change, and secondly, to propose a solution to automatic emotion detection using features that directly depict this change. The latter investigation is not directly related to the biometric problem, but the proposed analysis is very resourceful for real life deployment of identity recognition systems.

Human emotions are psycho-physiological experiences that affect all aspects of our daily lives. Emotions are complex processes comprised of numerous components including feelings, bodily changes, cognitive reactions, behavior and thoughts. Various models have been proposed by considering the ways in which these components interact to give rise

to emotions, but at the moment there isn't any single formulation that is universally acceptable. Modeling emotions is a very challenging problem (see Appendix B) that has drawn a great deal of interest from the emerging field of human-computer interaction (HCI). The objective is to design systems that can automatically identify emotional states, which would revolutionize applications in medicine, entertainment, education, safety etc. The main difficulty in formulating these models lies in the fact that we must rely on visible manifestations of emotions to produce, and verify them since the latent factors that generate emotions are unobservable.

The first step in modeling any phenomenon is data collection. We need to design experiments and institute methodologies that successfully induce emotions in a laboratory setting wherein we can record and collect psychological data. In quantifying psychological activity we are limited to the study of visible manifestations like facial expressions, gestures, vocal traits etc. These modalities are popular in HCI since they use the same cues that humans rely upon to detect and recognize emotional states. Moreover, most human beings display similar manifestations in response to identical emotional stimuli which allows for objective emotion annotation.

A major drawback of using behavioral modalities for emotion detection is the uncertainty that arises in the case of individuals who either are consciously regulating their emotional manifestations, or are naturally suppressive. For instance, although facial expressions can be analyzed to determine emotions, there is no guarantee that an individual will express the corresponding cue, irrespective of whether they are experiencing a certain emotion. This has serious implications in some applications such as surveillance.

Physiological signals, like the ECG, are an interesting alternative to the use of behavioral modalities. Other examples of this category include the electromyogram (EMG), electroencephalogram (EEG), galvanic skin response (GSR), blood volume pressure (BVP), heart rate (HR) or heart rate variability (HRV), temperature (T), respiration rate (RR). These signals have traditionally been used for clinical diagnostics, but there is signifi-

cant evidence to suggest that they are sensitive to, and may convey information about emotional states [86, 58, 69, 71, 80, 87]. One of the benefits of detecting emotions using physiological signals is that these are involuntary reactions of the body, and as such very difficult to mask. Moreover, for the duration of time that the sensors are attached to the body, these signals are recorded continuously, enabling frequent emotional assessment.

There are, however, many theoretical and practical challenges with regard to biosignal-based emotion detection. First, while the evidence suggests that physiological signals are affected by emotions, the exact effects on the waveform patterns remain to be seen. For example, the heart rate increases under both fear and excitement, but whether we can differentiate between the two is, as yet, unknown. Secondly, there are open questions about the subject specific nature of these effects.

Apart from the open theoretical questions, there are practical issues as well. The experimental protocols are far more complex than in behavioral emotion research, where the collection is facilitated by instructing volunteers to exhibit emotions. For biosignal-based experimentation more sophisticated practices are necessary to elicit truthful emotions in a laboratory setting. Furthermore, labeling physiological signals is subjective and as such very risky due to the difficulty in establishing the ground truth.

Another practical challenge relates to signal acquisition. The collection process is more invasive when compared to that for behavioral modalities, since the sensors need to be in contact with the human body for the duration of a recording session. For this reason it is important to minimize the amount of data required for this task i.e., to rely detection on as few signals as possible.

Different physiological signals originate from different locations of the human body and may describe unrelated functions. For instance, the ECG and BVP signals are of cardiovascular origin while the EMG relates to muscle electrical potential. It is important to investigate the dynamics of every signal in order to clearly establish its limits in assessing psychological activity. For instance, we anticipate that the experimental setup

alone may induce emotional reactivity for one biosignal, but may not stimulate some other.

To this end, this chapter investigates the feasibility and limits of emotion detection using ECG signals alone. Previous works have employed ECG (the HR mostly) in conjunction with other biosignals. There are several open questions such as understanding the psychophysiological rationale responsible for the formation of the signal, the subject specificity, the statistical limits to valence and arousal differentiation and the type of experiments that can stimulate it.

5.1.1 Signal Processing for Emotion Detection

There are two main challenges that arise with signal processing of physiological signals when targeting emotion pattern classification. First, emotion specific patterns are not well defined for biosignals. Secondly, it is typically uncertain whether emotions were manifested at all. For ECG signals, despite the reports on cardiovascular reaction to emotion, presented in Chapter 1, wherein there are inconclusive findings about the various waveform patterns, the majority of previous works consent on the subject specific nature of the emotion manifestation.

We argue that for ECG signals, universal emotion detectors run the risk of being inaccurate for the following two reasons:

1. ECG has been established as a biometric characteristic [88, 37], which means that by default it carries subject specific information. The appearance of a particular heart beat depends on a number of factors such as the geometry and orientation of the cardiac muscle, the conductivity of various areas of the heart, the activation order [17], and the subject's habitus or gender [11, 15, 18]. This *physiological aspect* of ECG formed a strong basis for the investigation and establishment of its biometric properties.

2. Different people experience emotions in different ways. Even if we ignore the biometric aspect of the ECG signal, there is large ANS innervation variability in a population [20].

For these reasons, emotional patterns are herein detected as variations from the typical appearance of an individual's ECG signal. We perform feature extraction in the time domain for two reasons – the time-domain signals have been reported to be more resourceful with regard to emotion specific features [89], and it is risky to impose stationarity and linearity restrictions (necessary for Fourier transform) on an ECG [90], especially because emotional activity operates on the signal in a non-linear fashion. Furthermore, since we cannot predict the effects of emotional processes on the ECG signal, there is a risk of missing the dynamic changes due to emotion if we use methodologies that rely on predefined bases.

We are more interested in the way properties of the ECG signal evolve over multiple heart-beats when a person experiences different emotions. The local properties of any particular pulse are of little interest. In this regard, Empirical Mode Decomposition (EMD) [90] is a powerful tool since it is dynamic, data-driven, and examines the signal holistically to highlight underlying trends. Without prior assumptions on the properties of the signal, EMD adapts to the embedded oscillatory activity and decomposes ECG into a number of intrinsic modes. It is our belief that understanding the intrinsic modes that are hidden in the cardiac oscillatory activity is an essential first step in any attempt to classify psychological states.

5.2 ECG-driven Empirical Mode Decomposition

The proposed methodology for ECG feature extraction is based on the Empirical Mode Decomposition (see Appendix A). EMD is adaptive and the basis of the decomposition is self-defined, which makes it suitable for the analysis of complex underlying phenomena.

However, it turns out that adaptivity is a mixed blessing. There are two important problems with regard to *uniqueness* and *mode mixing* that arise as a consequence of the adaptive nature of EMD. Evaluating the EMD algorithm on two instances of the same signal may result in incomparable decompositions. Moreover, each mode is not captured in a single decomposition - similar modes are present at various decomposition levels of the same signal. In our work, we address these issues by making use of a bivariate extension of EMD (BEMD) [91].

BEMD acts on two signals $x_I(t)$ and $x_S(t)$ (as opposed to EMD which acts on a single signal). In our work, the second signal ($x_S(t)$) is a synthetic ECG waveform, standardized over all emotional states and subjects, and designed to act as a decomposition guide.

The proposed framework is comprised of three independent steps:

1. ECG synthesis, wherein a signal $x_S(t)$ is designed to be synchronous with a particular input $x_I(t)$.

2. Estimation of the oscillatory modes, called *Intrinsic Mode Functions* (IMF) ($d(t)$), of the input signal via decomposition using the BEMD.

3. Extraction of features associated with the instantaneous frequency and the local oscillation of the IMFs and classification among predefined affect states.

5.2.1 ECG Synthesis

The objective of this step is to design a synthetic signal $x_S(t)$ whose properties are similar to a real ECG signal, and whose main waves are synchronized with the waves of the input signal $x_I(t)$.

In order to synchronize the main waves of the synthetic signal we must first estimate the location of the `P`, `QRS` and `T` waves throughout $x_I(t)$. The *QRS* complex is detected using the algorithm described in [92]. The surrounding waves are localized using empirical

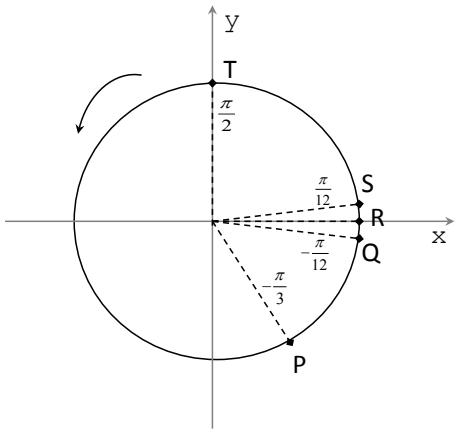


Figure 5.1: 2-D trajectory movement and `P Q R S`, `T` typical locations.

rules i.e., the `P` wave's healthy duration is approximately $120msec$, and the `T` wave extends about $300msec$ after the QRS complex [10].

Once the fiducial points of the input signal have been localized, the synthetic ECG signal can be generated using the dynamic generation model described by McSharry *et al.* [93]. The model generates a trajectory in a three-dimensional state space with coordinates `(x,y,z)`. Quasi-periodicity of the ECG is reflected by the movement of the trajectory around an attracting limit cycle of unit radius in the `(x,y)` plane, in varying speed. A completion of one cycle along this circle is equivalent to a heart beat completion or one `R-R` interval. By adjusting the speed of the trajectory, quasi-periodicity is achieved. Inter-beat variation in the ECG signal is reproduced by the motion of the trajectory in the `z` direction.

The location of the fiducial points of the input signal define special events on the unit circle indexed by the angles θ_P, θ_Q, θ_R, θ_S, θ_T corresponding to points P, Q, R, S, T respectively. These distinct points on the ECG are described in the model by events corresponding to the negative and positive attractors/repellors in the `z` direction. Each time the trajectory reaches one such point, it is repelled away from the unit circle along the `z` axis and then attracted back, creating a wave. The dynamical equations of motion

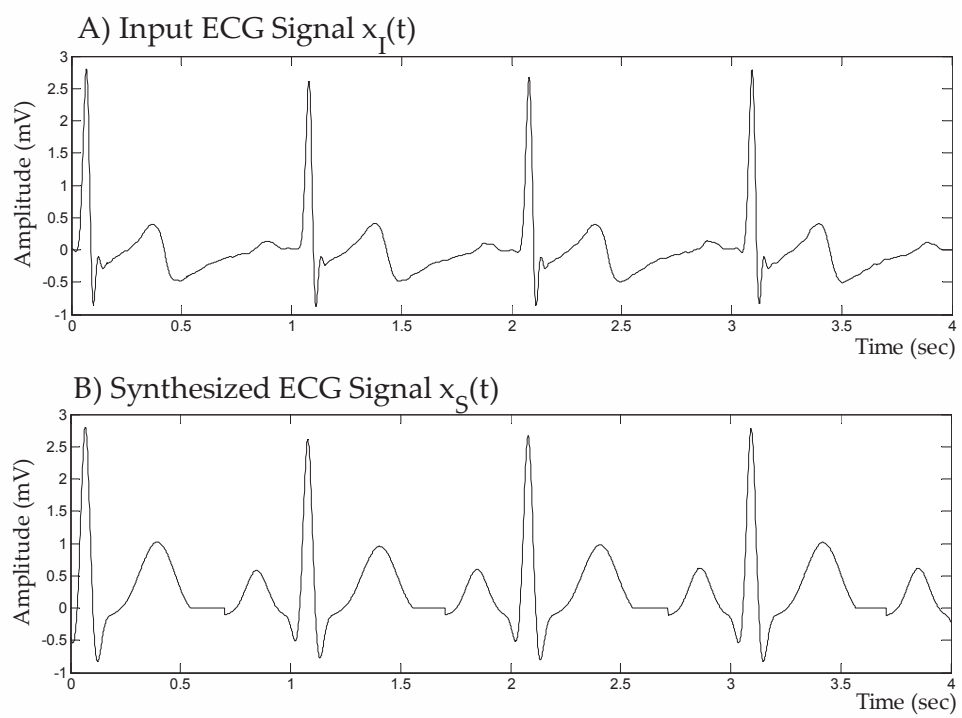


Figure 5.2: A real and a synthetic ECG. The two signals are synchronized but $x_S(t)$ has no anatomical uniqueness or psychological variability.

along this trajectory are given by a set of three ordinary differential equations [93].

The authors in [93] provided typical values for the angular location of the fiducial points (Fig. 5.1) based on visual inspection of healthy ECG waveforms. The duration of one circle is set to a maximum of one second (rest condition, 60 bpm). Given the distances d_{ij} (in seconds) between the fiducial points i and j of $x_I(t)$, the points can be located along the unit circle using the following equations (wherein the location of a point is specified in radians):

$$\theta_P = -2\pi \frac{d_{PR}}{d_{RR'}} \tag{5.1}$$

$$\theta_Q = -2\pi \frac{d_{QR}}{d_{RR'}} \tag{5.2}$$

$$\theta_S = 2\pi \frac{d_{RS}}{d_{RR'}} \tag{5.3}$$

$$\theta_T = 2\pi \frac{d_{RT}}{d_{RR'}} \tag{5.4}$$

where $d_{RR'}$ is the length of the R-R interval in seconds, and $\theta_R = 0$. These angles are applied on the differential equations in [93] to obtain $x_S(t)$. The generated signal is an idealized, robust, noise free representation of ECG (see Fig. 5.2 for an example of a synthetic ECG signal).

5.2.2 Signal Decomposition

Huang *et al.* [90] proposed Empirical Mode Decomposition as a way to empirically decompose a non-stationary, non-linear signal into a number of IMFs, each of which represents a distinct oscillatory activity. An IMF is a function satisfying certain explicit properties [90]:

1. The number of extrema and the number of zero crossings must be equal or differ at most by one.

2. The mean of the envelopes defined by the maxima and the minima is zero for every sample (see Appendix A).

These rules naturally force one mode of oscillatory activity in the IMF, since between two successive extrema no riding waves are allowed. The algorithm for the detection and extraction of IMFs is adaptive and iterative. Once an IMF is found, it is removed from the signal, and the algorithm iterates on the residual in order to find more oscillatory modes. Fast oscillations are detected first. The basic EMD algorithm along with an example ECG signal decomposition is provided in Appendix A.

Overall, the decomposition is given by:

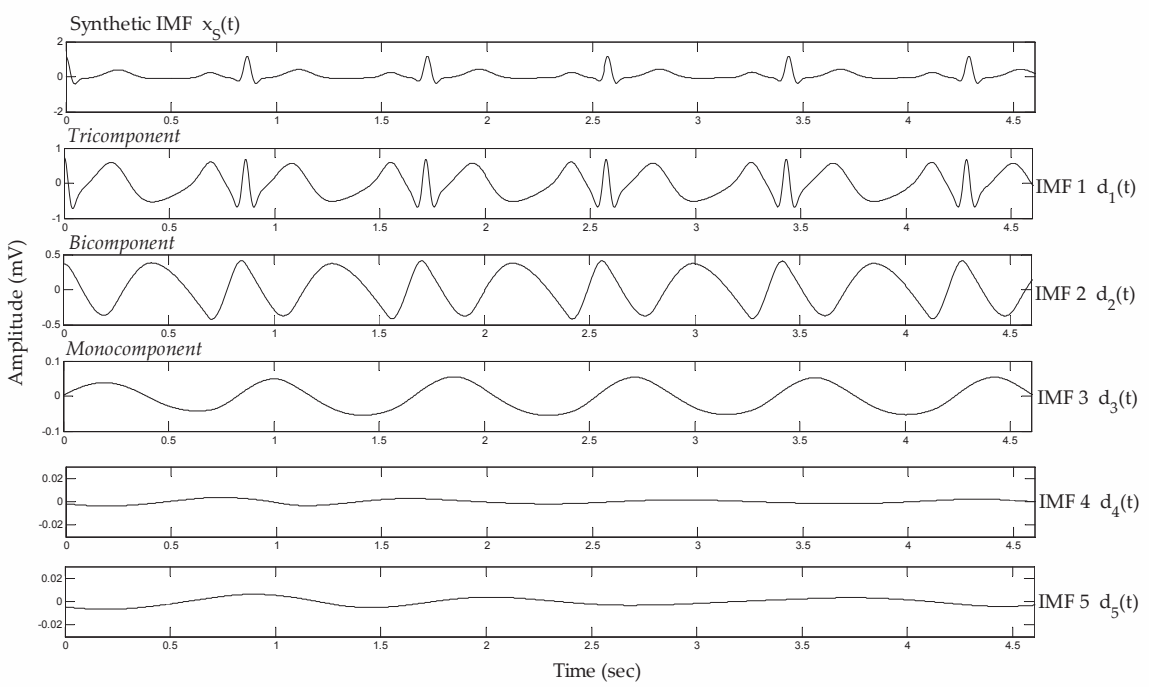


Figure 5.3: EMD analysis for a standardized synthetic ECG signal. IMFs of order higher than three do not exhibit oscillatory activity. The first IMF has three oscillatory components, the second has two and the third has one.

$$x(t) = \sum_{i=1}^{N-1} d_i(t) + r(t) \tag{5.5}$$

where $d_i(t)$ denotes the i^{th} IMF extracted from the signal $x(t)$ and $r(t)$ is the final residual.

Figure 5.3 shows an example of the resulting IMFs when EMD is applied on a synthetic signal $x_S(t)$ which is an idealized, noise-free waveform. It has been observed [94, 89, 95, 96] that the first few IMFs carry the quasi-periodicity property of ECG. Since every heart beat has three distinct waves, in the absence of noise, the first IMF is expected to exhibit three oscillatory components (*tricomponent*), characterizing primarily the behavior of the `QRS` complex. This is because the `QRS` complex contributes to the highest frequencies of the ECG. The first IMF, as shown in Figure 5.3, depicts the fastest oscillating component of the signal. Once this is removed, the second and third IMFs exhibit *bicomponent* and *monocompoent* oscillations respectively. As the IMF order increases, the strength of the

oscillation decreases. However, for noise free ECG, IMFs of order higher than three are almost zero, and will be ignored hereafter.

As stated before, the EMD encounters problems with regard to *uniqueness*. The number and type of IMFs generated by EMD are uncertain, even for signals with similar statistics. For instance, the decomposition in Figure 5.3 would result in more IMFs and in stronger oscillations if there was high frequency noise in the signal. This restricts the utility of EMD as it renders comparisons among different (but real) ECG signals meaningless. Predetermining the number of IMFs (by forcing decomposition to stop) defeats the purpose of EMD as the analysis will no longer be adaptive, nor will the IMFs have physical meaning.

We address this problem by decomposing a real ECG signal $x_I(t)$ together with a synthetic signal $x_S(t)$ in a module that allows the latter to act as the *rule* of decomposition by determining which type of IMFs are important from the input signal. This is done using the BEMD [91] algorithm on the pair of signals $x_I(t)$ and $x_S(t)$. The difference between EMD and BEMD can be visualized as follows: whereas EMD-sifting builds envelopes around $x(t)$, BEMD builds 3D cubes that surround a complex function $x_c(t)$. The analysis is performed simultaneously for the real and imaginary components of $x_c(t)$, and results in the same number of IMFs for both:

$$x_c(t) = \sum_{i=1}^{N-1} d_{c_i}(t) + r_c(t) \tag{5.6}$$

where $d_{c_i}(t)$ denotes a complex IMF and $r_c(t)$ the complex residual. Low order $d_{c_i}(t)$ describe fast rotating components, while the opposite is the case for higher order complex IMFs. Because of the consistency in the analysis of the real and the imaginary parts, BEMD has been suggested for signal separation (de-trending) in filtering applications [97, 98]. For our purpose, we form a complex signal using the input ECG signal $x_I(t)$ and the synthetic signal $x_S(t)$ as the real and imaginary parts respectively.

$$x_c(t) = x_I(t) + jx_S(t) \tag{5.7}$$

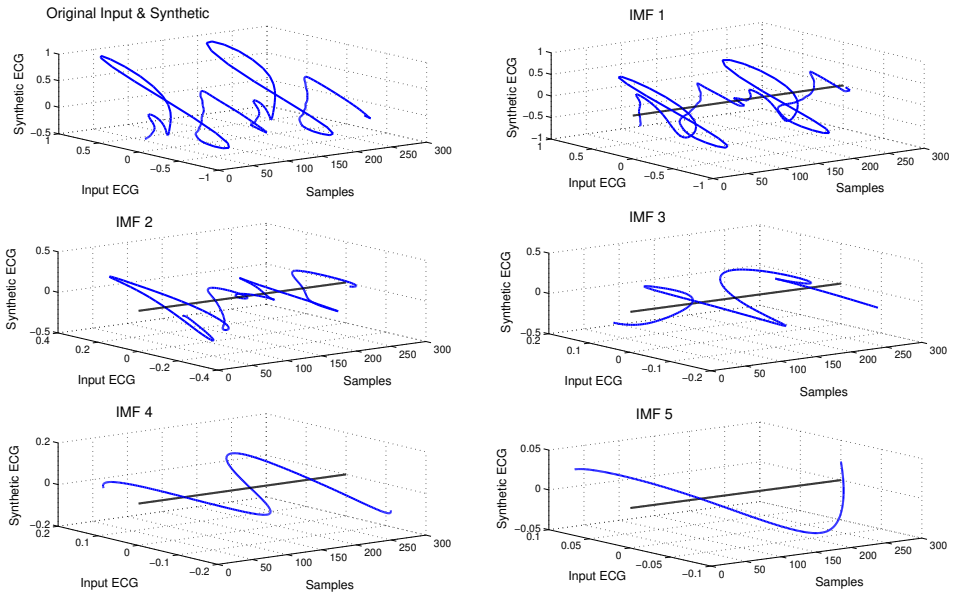


Figure 5.4: BEMD example on a complex ECG signal, formed using a real ECG segment and a synthetic one. Low order IMFs show fast rotating components.

Applying BEMD on $x_c(t)$ we get

$$x_c(t) = \sum_{i=1}^{N} Re\{d_i(t)\} + j \sum_{i=1}^{N} Im\{d_i(t)\} \tag{5.8}$$

where the residual has been included in the summation for simplicity. Similar to the univariate case, Figure 5.4 shows an example of five IMFs acquired from a BEMD analysis of two heart beats (one real and one synthetic). Similarly, Figure 5.5, shows the simultaneous decomposition of the two signals for the real and imaginary parts separately. Since the synthetic ECG has an idealized waveform, the presence of oscillatory activity on the imaginary side guarantees that the corresponding mode is present on the real side. When the real ECG is contaminated with high frequency noise, low order IMFs on the real part will exhibit strong (but physiologically meaningless) oscillations while almost zero activity will exist on the imaginary side which makes them easily detectable.

Figure 5.6 shows a comparison between the univariate EMD and the driven BEMD decompositions for the same ECG segment. Even though there is no theoretical guaran-

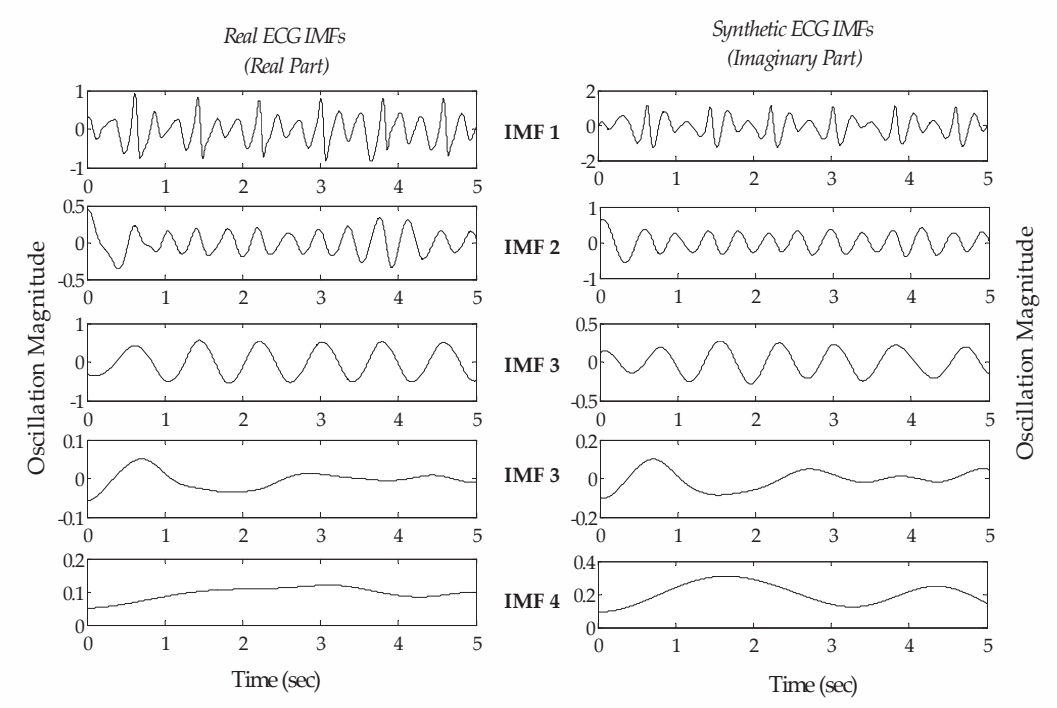


Figure 5.5: Simultaneous decomposition of real and a synthetic ECG signal using the BEMD.

tee for uniqueness, the proposed driven BEMD algorithm bypasses this inadequacy by ensuring that the three most substantial IMFs for ECG analysis will be present in the decomposition without mode mixing.

5.2.3 Feature Extraction

The IMFs are time domain signals carrying information about oscillation activity. Comparisons in the time domain are not straightforward since the IMFs $d_i(t)$ have to be aligned with similar modes from other ECG recordings. Therefore, it is important to design features that summarize the oscillatory activity within every IMF. In this work, we use two types of features i.e., the *Hilbert instantaneous frequency* and a measure of *local oscillation*.

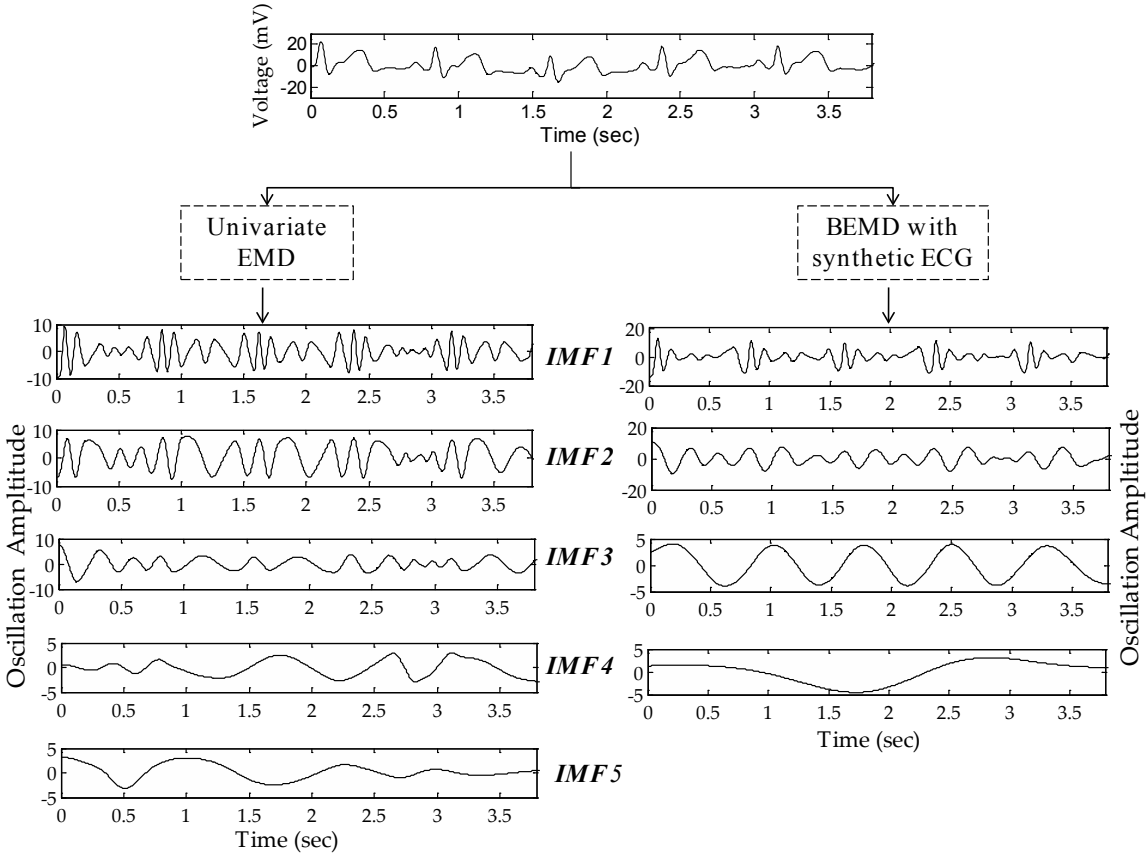


Figure 5.6: Comparison of Univariate and driven Bivariate EMD decomposition on the same ECG signal. For the BEMD case the IMFs exhibit less mode mixing as well as the oscillation structure follows the properties of ECG decomposition in the absence of noise i.e., IMF 1 is tricomponent, IMF 2 is bicomponent and IMF 3 is monocomponent.

5.2.3.1 Instantaneous Frequency

The Hilbert transform is typically used in conjunction with the EMD because it provides accurate estimates of the instantaneous frequency for monochromatic signals like the IMFs. The transform is defined as the convolution of a signal with $h(t) = \frac{1}{\pi t}$. For each of the IMFs $d_i(t)$, the Hilbert transform is applied as follows:

$$\mathcal{H}[d_i(t)] = \frac{1}{\pi} P.V. \int_{-\infty}^{+\infty} \frac{d_i(\tau)}{t-\tau} \, \mathrm{d}\tau \tag{5.9}$$

where P.V. indicates the Cauchy principal value. We can define the following analytical signals:

$$z_i(t) = d_i(t) + j\mathcal{H}[d_i(t)] \tag{5.10}$$

which can be rewritten as

$$z_i(t) = y_i(t)e^{j\theta_i(t)} \tag{5.11}$$

where $y_i(t)$ is the magnitude and $\theta_i(t) = \arctan\left(\frac{\mathcal{H}[d_i(t)]}{d_i(t)}\right)$ is the phase of the complex $z_i(t)$. For IMF i the instantaneous frequency can then be computed for as:

$$f_i(t) = \frac{1}{2\pi} \frac{\mathrm{d}\theta_i(t)}{\mathrm{d}t} \tag{5.12}$$

Essentially, $f_i(t)$ is a measure of changeability within every IMF.

5.2.3.2 Local Oscillation

Modeling the type of oscillation within an IMF is a difficult problem due to the empirical nature of EMD. We herein propose features that are extracted as measures of the oscillation time-scale, rather than the velocity or frequency of alternation among extrema. For every IMF $d_i(t)$, let $\mathbf{u}_i$ and $\mathbf{v}_i$ be the time instances of the maxima and minima respectively. By definition (of IMFs), there is one zero crossing between every pair of consecutive extrema and the objective is to define the rate of interchange.

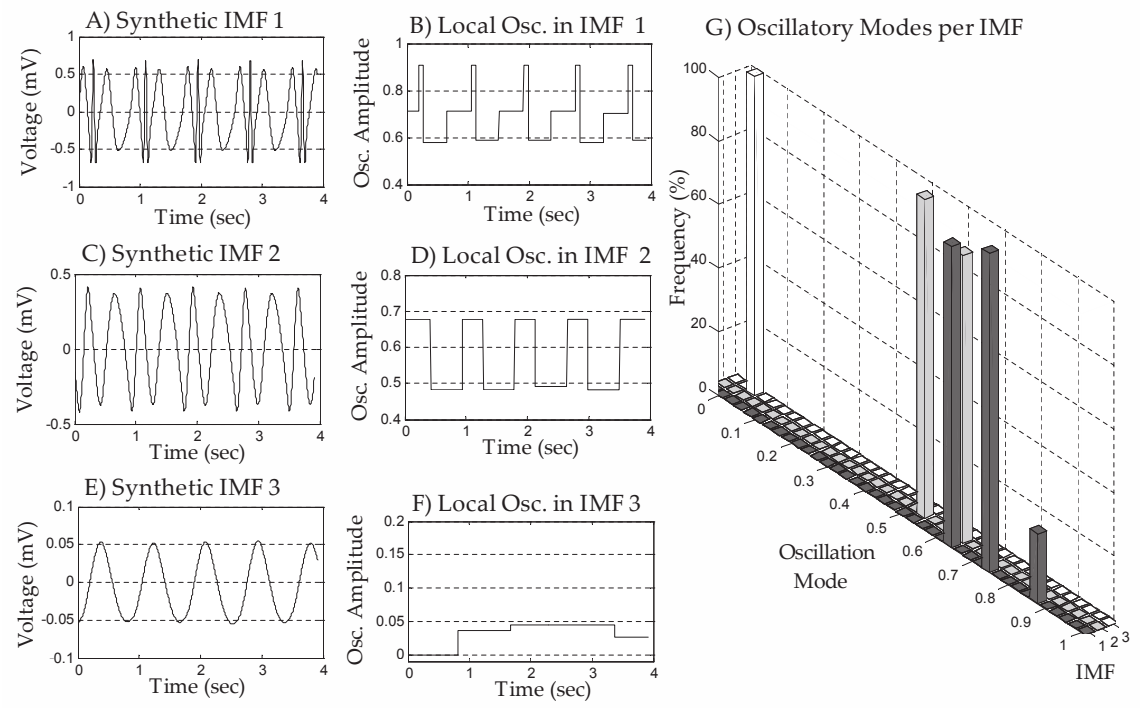


Figure 5.7: Local oscillation for a synthetic ECG. A,C,E) First three IMFs of the synthetic signal B,D,F) $\rho_i(t)$ for the previous IMFs G) Dominant oscillations for the three IMFs. With increasing order of IMF, the strength of the oscillation (time-scale) decreases.

We define $\rho_i(t)$, a function that describes local oscillation, which is estimated by parsing an IMF as follows:

1. For an element $\mathbf{u}_i^k$ of $\mathbf{u}_i$, examine $d_i(t)$ in the interval $[d_i(u_i^k), d_i(u_i^{k+1})]$ (between two consecutive maxima).

2. Compute *max-to-min* and *min-to-max* transition times: $a = v_i^k - u_i^k$, $b = u_i^{k+1} - v_i^k$

3. $y_i(t) = \min(a, b)$, $u_i^k \leq t \leq u_i^{k+1}$.

The local oscillation $\rho_i(t)$ is then computed from $y_i(t)$ with normalization across all IMFs as follows:

$$\rho_i(t) = \frac{1 - y_i(t)}{A} \tag{5.13}$$

where A is the maximum of all $\rho_i(t)$ i.e., local oscillation is normalized across all IMFs. The higher the values in $\rho_i(t)$, the faster the local extrema interchange. Figure 5.7 shows $\rho_i(t)$ for the first three IMFs of a synthetic ECG. The oscillation histogram shows that the first IMF has three major components, the second has two and the third just one. For a measured ECG signal, this distinction is unclear, as mode mixing is a very common phenomenon for EMD i.e., different types of oscillatory patterns appear throughout one IMF.

To address this problem, classification is carried out only among the *dominant* local oscillations and frequencies. The feature vector is the concatenation of these measures for the first three IMFs, which exhibit quasi-periodicity. More specific, since the first IMF is a tricomponent signal, the three most frequent local oscillations, referred to as $\hat{\rho}_1^1, \hat{\rho}_1^2$ and $\hat{\rho}_1^3$ (see Fig. 5.7.G), will participate in the feature vector. Similarly, three frequency measures are considered for the first IMF ($\hat{f}_1^1, \hat{f}_1^2$ and $\hat{f}_1^3$). Overall, the feature vector is formed as follows:

$$x = [\hat{\rho}_1^1, \hat{\rho}_1^2, \hat{\rho}_1^3, \hat{\rho}_2^1, \hat{\rho}_2^2, \hat{\rho}_3^1, \hat{f}_1^1, \hat{f}_1^2, \hat{f}_1^3, \hat{f}_2^1, \hat{f}_2^2, \hat{f}_3^1] \tag{5.14}$$

Classification is carried out using linear discriminants among a number of predetermined classes. The assumption in this work, to be verified experimentally, is that both the local oscillations and the instantaneous frequencies carry emotion specific information.

5.3 Performance Evaluation

The proposed methodology was evaluated on signals from the Passive and Active Arousal ECG Databases (Chapter 3). In the first experiment, visual stimuli (passive stressors) were used to induce positive and negative emotions for valence and passive arousal performance evaluation. The second experiment attempted to induce active mental stress (*active arousal*), using a video game.

<table>
<tr><td>5-class: Distinct emotions</td><td colspan="2">Experiment A class 1: Erotica, class 2: Excitement, class 3: Fear
class 4: Disgust, class 5: Gore</td></tr>
<tr><td rowspan="3">Valence Differentiation</td><td rowspan="2">Arousal Dependent</td><td>Experiment B class 1: Erotica class 2: Gore</td></tr>
<tr><td>Experiment C class 1: Erotica class 2: Disgust</td></tr>
<tr><td>Arousal Independent</td><td>Experiment D class 1: Erotica, Excitement
class 2: Fear, Gore, Disgust</td></tr>
<tr><td rowspan="4">Within Valence Differentiation</td><td>Within Positive</td><td>Experiment E class 1: Erotica class 2: Excitement</td></tr>
<tr><td rowspan="3">Within Negative</td><td>Experiment F class 1: Fear class 2: Disgust</td></tr>
<tr><td>Experiment G class 1: Gore class 2: Disgust</td></tr>
<tr><td>Experiment H class 1: Gore class 2: Fear</td></tr>
<tr><td>Arousal Differentiation</td><td colspan="2">Experiment I class 1: Erotica, Gore class 2: Neutral</td></tr>
</table>

Table 5.1: List of classification experiments performed.

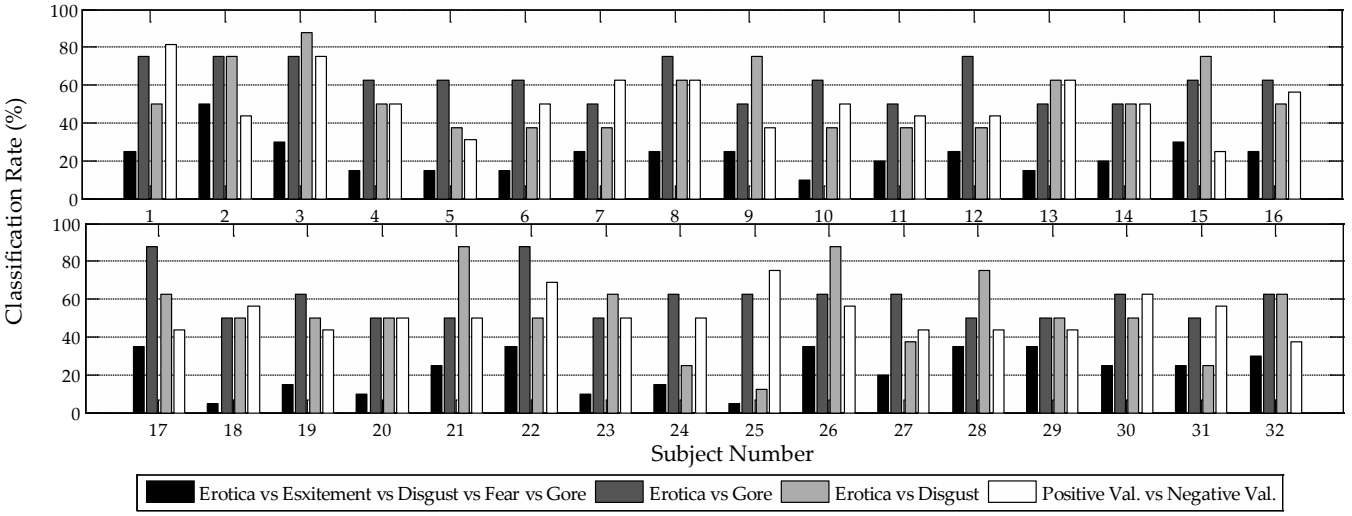


Figure 5.8: Per subject classification performance for each of the 32 individuals in the database. While classification among five classes is promising for certain individuals, valence separation is feasible with respect or irrespective of arousal (erotica vs gore or erotica vs disgust) .

5.3.1 Valence classification

A first step to ECG signal processing is noise filtering. In this setup, a butterworth bandpass filter with cutoffs at 0.5Hz and 40Hz was used. The order of the filter was set to 4 based on empirical results. Further processing of the signal is slightly different for the two experiments. In the IAPS case, the signal is divided into segments corresponding to the particular emotional conditions of the experiment (erotica, excitement, disgust, fear, gore, neutral). Every segment was further subdivided into a number of windows of 10 sec length. Practically, the goal was to classify a state every 10 sec.

Out of the examined conditions, erotica and excitement fall under positive valence, with erotica being of higher arousal. On the negative valence side, gore exhibits the strongest arousal. A variety of experiments were conducted as listed in Table 5.1. The goal was to demonstrate that ECG can differentiate among or within the same valence conditions, when this operation is subject dependent.

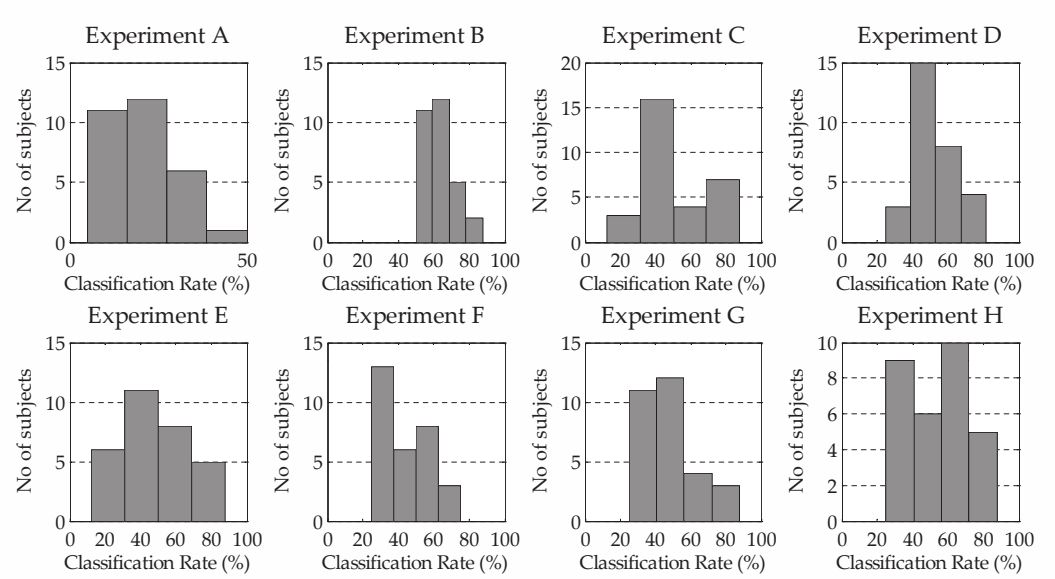


Figure 5.9: Classification performance for all subjects in experiments A-H.

Synthetic ECG signals where generated and participated in BEMD analysis for each of the 10 sec segments. The instantaneous frequency and the local oscillations of the first three IMFs were used as classification features. Since pictures of the same condition were displayed twice, data from the first batch were used for training and from the second one for testing. Classification was performed with linear discriminants, and with leave one out cross validation.

Figure 5.8 shows per subject performance for experiments A, B, C and D. It can be observed that among highly aroused conditions (gore and erotica), valence can be differentiated for most subjects, but not for all. For certain individuals erotica is more distinguishable from disgust than gore. As it will be clarified in the subsequent discussion, this observation is related to the intensity with which every participant perceived the experiment. For individuals that reported stronger immersion, experiment C indicates higher classification performance.

The performance for all experiments in Table 5.1 is graphically depicted in Figure 5.9. It is important to note that for each case there are individuals which score considerably

high, but it is difficult to generalize especially for the within valence classification cases. However, valence differentiation (in experiment B) is clearly feasible as performance ranges from *52%* to *89%* with a standard deviation of *11.16*. An interesting finding is also with regard to classification between two negative valence but high arousal states (gore and fear). As it can be seen from Figure 5.9 ECG can differentiate within negative emotions with higher probability than within positive. However, this might be attributed to the fact that negative emotions were better induced with visual stimuli.

5.3.2 Arousal classification

For the data of the active arousal experiment, a similar windowing approach was adopted (10 sec segments). Since the self assessment was done using FEELTRACE [81], labeling was continuous for the duration of the video game. On FEELTRACE, arousal takes values in $[-10, +10]$ with -10 indicating very low emotional intensity (boredom) and $+10$ the opposite. Depending on the desired precision, arousal was quantized to a number of levels, and the average within every 10 sec window was used to characterized it.

Two classification problems were examined i.e., a three class problem of discrimination among *low* ($[-10, -6]$), *medium* $(-6, +6)$) and *high* arousal $[+6, +10]$), and a two class problem of discrimination between *low* $[-10, 0)$) and *high* $[0, +10]$) arousal.

Because for a game experience it is difficult to report instantaneous valence information, the volunteers were asked to focus their reporting to the arousal dimension. However, continuous self reporting is ambiguous first, because determining the actual onset of excitement or boredom can be misleading, and second, because even if the onset of a cognitive reaction is accurately determined, there might be cardiac latency especially for transitions between high-to-low arousal. For example, when an interesting event in the game is followed by a monotonous situation (for instance because the subject got lost in the map), the progression between high-to-low arousal is gradual and it is thus difficult for the player to assess the exact onset of boredom. Perfect data annotation is

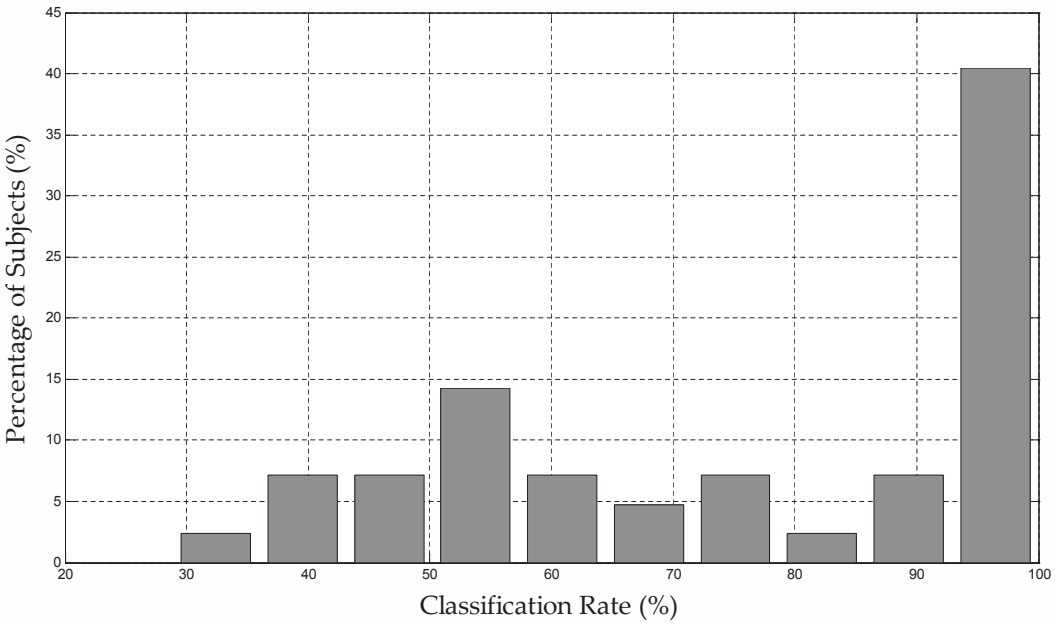


Figure 5.10: Active arousal detection performance for 42 subjects.

of course of great importance, and this was expected to add to the overall classification error. However, the immersion that video games provide, in conjunction with the user being actively engaged to the task (active stressor) proved beneficial for the problem in hand.

For every subject, the first 10 min of data were used for training and the remaining for testing. Every feature vector was the concatenation of instantaneous frequency and local oscillation measures. Classification was performed with linear discriminants, and it was subject dependent.

For the three class problem (*low*, *medium* and *high* arousal), the average classification rate over all subjects is *35.36%*. For the two class problem (*low* and *high* arousal), the average rate over all subjects is *76.19%*. The mean however is not very descriptive for the population of 42 individuals, were classification is subject dependent. Figure 5.10 shows the probability mass function for the two class problem. For approximately *50%* of the subjects, a detection rate between *90%-100%* was achieved.

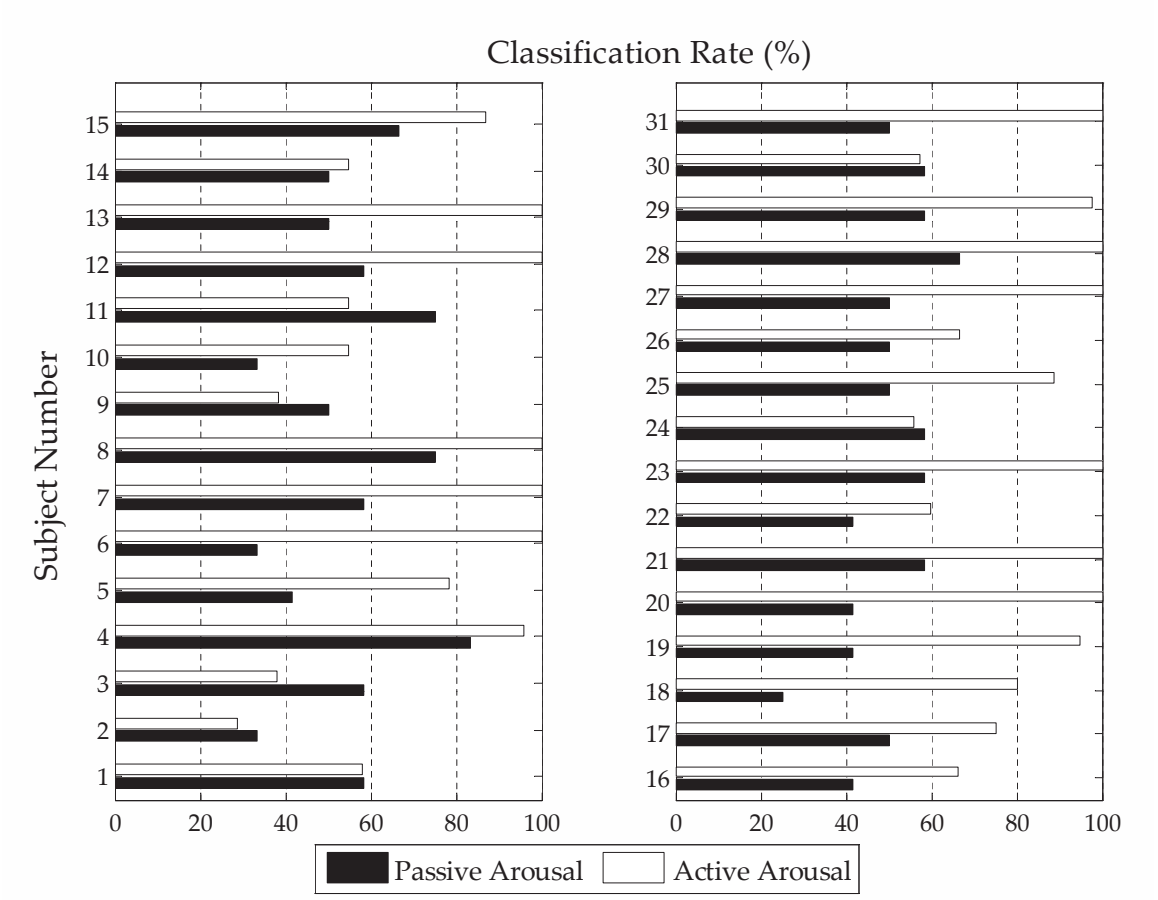


Figure 5.11: Subject specific arousal detection for the two experiments. The average rate for all subjects under passive arousal is *52.41%*. Similarly for active arousal the average is *78.43%*.

5.3.3 Active vs passive arousal

Inspired by the reports of varying cardiac reactivity due to passive or active stressors [31, 32, 99], a performance comparison between arousal classification in experiments I and II was performed. In total there are 31 subjects who participated in both experiments, and for which artifact free data are available. Performances were compared for the two class problems of discrimination between *high* and *low* arousal. In experiment I this was achieved by distinguishing neutral reactions from the union of gore and erotica conditions.

Figure 5.11 lists the classification performance per subject, for both passive (exper-

iment I) and active (experiment II) arousal. The average rate is *52.41%* and *78.43%* respectively. For the majority of the subjects active arousal detection outperforms passive. For the few subjects that this is not the case, the reason might be that visual stimuli were of equal immersion as the game.

5.4 Chapter Summary

In the emotion research it is very important to collect meaningful affect data. However, it is very difficult to design an experimental setup that can induce the same emotion in every subject, especially if the same stimuli are used across all subjects. Different characters, varying moods and the inability to accurately self report an emotional experience, may significantly affect the outcome of a study. In particular, when internal modalities like the ECG are examined, data labeling is very subjective and can only be verified by the participants themselves.

Despite the difficulties, establishing emotions from internal manifestations of the body is worthwhile for human computer interaction systems. For the naive user, hiding emotions with regard to cardiac reactions is difficult, while for behaviorally suppressive individuals physiological patterns can provide hints of emotion. The obtrusive nature of the acquisition however, poses restrictions on the number of sensors that can be worn and subsequently to the signals that can be collected. The majority of the approaches in the literature rely on fusion of various physiological sources for emotion detection. Although such treatment presents significant performance benefits, they have limited practicality. Following the findings of this study, future treatments will attempt to establish the limits of emotion assessment for other biosignals as well.

Although ECG has been extensively employed in the affect research, little attention has been paid to the waveform patterns under emotional activity. A typical approach is to extract the R-R intervals for the computation of the HRV, and treat the remaining signal

as redundant. However, given that the ANS has endings in each of the four chambers of the heart, it is expected that ECG will exhibit emotion specific patterns. However, for any analysis to meaningful one has to take into consideration the specificity of the signal to the particular subjects due to both the biometric aspect of the signal and the specificity of the ANS innervation.

This chapter proposed subject dependent ECG signal decomposition using the BEMD. Local oscillation and instantaneous frequency features were used for the detection of emotional conditions, as their combination describes fully the oscillatory activity of every IMF. It was concluded that adaptive, data-driven analyses are suitable for emotion modeling, because of the unforeseeable patterns that may arise within every individual. With the employment of oscillation data as the feature space, it was shown that ECG waveform reactivity depends highly on the activeness of the emotional experience. Data from visual stimuli inspection resulted in *26.02%* lower arousal detection performance (on average for 31 volunteers) than playing a video game. This finding is in agreement with prior reports [31, 32, 99] on cardiovascular differentiation between active and passive stress. Valence could also be differentiated, especially for high arousal states with an average of *62%* and a standard deviation of *11.16*.

This chapter statistically demonstrates that ECG carries emotion specific information. It was shown that emotional activity is detectable even in cases where the heart rate is not affected. From a biometric recognition point of view, the fact that active arousal may alter the signal is critical. In most real life situations, emotions are experienced while the individual is engaged in tasks such as working, problem solving, teaching, speaking or simply cooking. While passive situations, such as watching a movie or listening to music, do not threaten the stability of the signal, it is important that the destabilization of the biometric template under active arousal is taken into account. The following chapter provides the means for early detection of destabilization and template update.

Chapter 6

Continuous Authentication

6.1 Problem Statement

Section 4.4.4 of Chapter 3 emphasized the shortcomings of the ECG signal's time dependency by demonstrating how a biometric template can destabilize with time. Moreover, Chapter 4 discussed and experimentally validated the conditions under which human emotion may be detected from ECG signals.

These two factors motivate a special treatment that can rectify the weakness due to time-dependency. Practically, there are two ways to address this problem. One option is to look for features that are robust to psychological activity. The problem with this approach is that isolation of robust features would require an explicit definition of ECG morphologies under different emotions, which is difficult to acquire. The reader should note that the oscillatory features that were proposed in Chapter 4 are subject dependent and there is no straightforward way of separating them from the rest of the waveform.

The second option is to automatically detect signal destabilization with respect to a template, and update the template if necessary. This signal destabilization, or decoherence, is caused by changes in psychological states. In this case, the system is interested in detecting state changes (i.e., detecting coherence intervals) rather than classifying emo-

tional states. This solution can be directly applied to monitoring environments where the ECG signal is input continuously to the system. In this chapter we explore the second option.

6.2 Template Updating

The objective of the subsequent analysis is to update the biometric template at instances corresponding to the destabilization or decoherence of the correlation values as discussed briefly in Section 4.4.4 (the perils of ignoring template destabilization are graphically demonstrated in Figure 4.13).

Since the purpose is to detect state changes, and not to classify emotions, a coherence analysis approach is herein adopted. The idea is to design variable-length accumulated durations based on some *fundamental* time duration. Consider, for instance, the various coherence durations shown in Figure 4.11. It is easy to see that these durations are dependent not only on the subject, but also on the psychological activity at the time of measurement. Therefore, it is expected that a fixed updating duration may not be efficient. This intuition is experimentally validated, and the results are shown in 6.3.2.

There is a trade-off between frequency of template-updates and computational complexity. Frequent template updating implies accurate tracking of events, but increases the computational complexity. On the other hand, infrequent template updating may cause inadequate system performance in terms of increased false rejection.

Taking these considerations into account, the idea of defining a fundamental time duration (which is subject or application dependent) is to consider some acceptable *minimum duration*, over which the system makes a decision about whether the template needs to be updated. The extreme case is to define this fundamental duration to be equal to the smallest time-resolution in the system (5 seconds for the AC/LDA algorithm). However, this is computationally inefficient and therefore the updating instances

need to be strategically chosen.

In essence, a variable-length accumulated duration (or *burst*) is constructed by accumulating various fundamental durations (for the proposed AC/LDA algorithm, these fundamental durations correspond to ECG segments of 5 seconds length). The following iterative description can be made.

Consider the following fundamental durations $\{d_1, d_2, \ldots\}$, where each d_i corresponds to time duration of 5 seconds. This duration is chosen to acceptably accommodate the time resolution requirement of the AC/LDA algoritm. Now suppose that at the current iteration, the current burst,$\mathbf{D}_{\text{current}}$, contains μ fundamental durations, i.e., $\mathbf{D}_{\text{current}} = \{d_k, \ldots, d_{k+\mu-1}\}$. For the subsequent segment, $d_{k+\mu}$, the two choices are:

(C1) Add $d_{k+\mu}$ to the current burst, forming $\mathbf{D}_{\text{potential}} = \{d_k, \ldots, d_{k+\mu}\}$. Continue the operation with $d_{k+\mu+1}$ as the next candidate.

(C2) Reject $d_{k+\mu}$, and terminate $\mathbf{D}_{\text{current}}$. Re-initialize with $d_{k+\mu}$ as the start of a new burst.

When the correlation coefficient profiles (as in Figure 4.11) are utilized to decide between (C1) or (C2), the following procedure is performed:

1. Compute the correlation profile for $\mathbf{D}_{\text{potential}}$ relative to starting point d_k.

2. Find the minimum correlation value $c_{\min}$ over $\mathbf{D}_{\text{potential}}$.

3. Compare to a threshold c_{th} for decision:

$$c_{\min} - c_{th} \underset{C_2}{\overset{C_1}{\gtrless}} 0. \tag{6.1}$$

The algorithm described above is summarized in Table 6.1.

It should be noted that the described algorithm includes a parameter $dsize_{max}$, which is used to control the maximum size of the burst, i.e., to force template updating. Such a strategy could be necessary to deal with buffering requirements (since the entire accumulated data needs to be stored), as well as to reset the algorithm in case of misdetections.

c_{th}: threshold for decision

N_{total} : total number of fundamental durations to be processed

$dsize_{max}$: max number of fundamental durations accumulated

s: fundamental duration defining start of the current accumulated duration

I. ***Initialization***

1. Set $s = 1$

II. ***Iteration***

for $i = 2, 3, \cdots, N_{\text{total}}$

if $(i - s + 1 \geq dsize_{max})$ **or** $(i == N_{\text{total}})$

1. Set current burst =

all fundamental durations from s to i

2. Update the biometric template

3. Reset $s = i + 1$

else if $(c_{min} < c_{th})$

1. Set current burst =

all fundamental durations from s to $i - 1$

2. Update the biometric template

3. Reset $s = i$

end

end

Table 6.1: Template updating with variable-length durations

6.3 Performance Evaluation

The performance of the proposed template-updating scheme is evaluated over ECG recordings with and without affective labeling. First, the effects of frequency of updates

are assessed over the 10 individuals included in the long-term ECG database presented in Section 4.4. This analysis demonstrates the benefits of frequent template updating, without validation on the psychological status of the subjects. A similar analysis is then presented for the active arousal database (Chapter 3). For the latter database we had showed that emotional states can be successfully classified, from which we can infer that there is great risk of template destabilization.

6.3.1 Effect of Template Updating Frequency on System Performance

To numerically quantify the effects of frequency of template updating on performance, over the long-term ECG dataset, three scenarios are considered in measuring recognition performance namely, *high*, *moderate* and *low* correlation between the biometric template and the testing ECG segments. In essence, this was simulated by adjusting the correlation threshold c_{th} which guides the updating procedure. The lower this threshold the more the variability allowed within the burst, and therefore the smaller the frequency of the updates. Accordingly, higher thresholds pose great coherence restrictions and require more frequent updates. More specific, the following cases were investigated:

1. **Highly correlated training and testing windows (found close in time).** This scenario effectively represents the lower-bound in error rates corresponding to frequent template updating because of high c_{th} restrictions. Due to frequent template updating, the bursts constructed are characterized with very high correlation. However, this case may not be desirable from a computational complexity perspective. Figure 6.1 illustrates the error and authentication rates for this case. The EER for frequently updating the template is *3.4%*.

2. **Moderate correlation scenario.** This corresponds to the use of template updating with a moderate c_{th} selection. In this case, some bursts may suffer from misde-

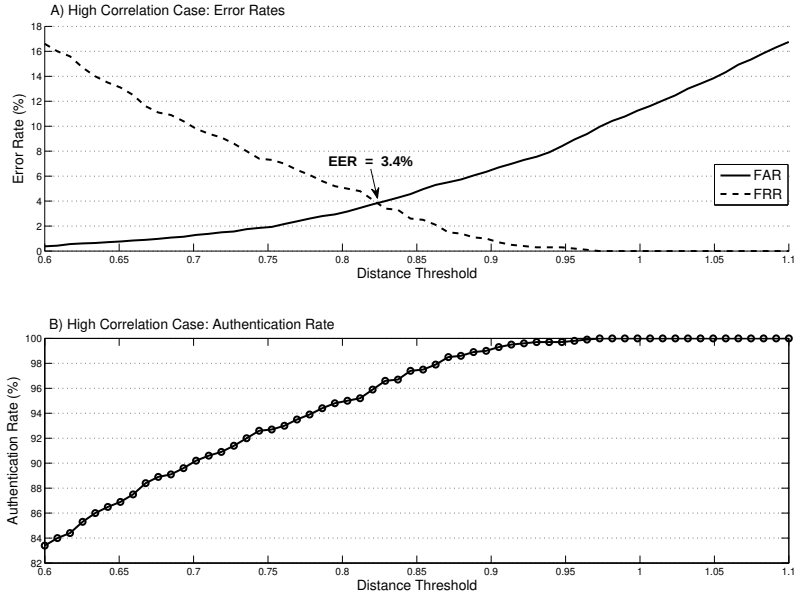


Figure 6.1: Verification performance for *highly* correlated training and testing ECG windows.

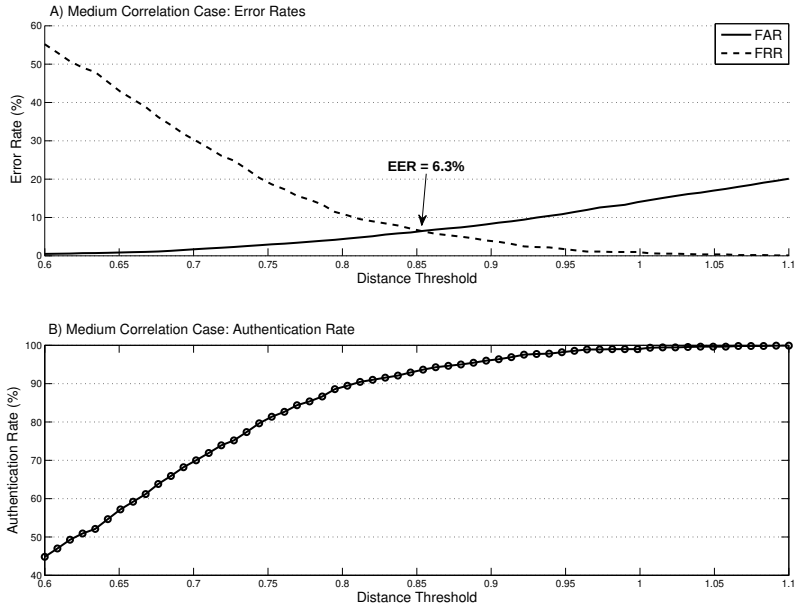


Figure 6.2: Verification performance for *moderate* correlation between the training and testing set.

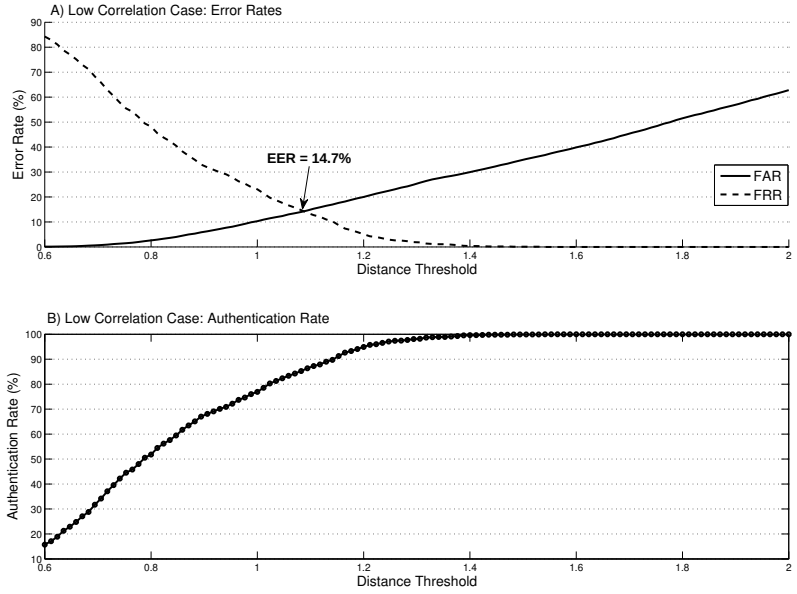


Figure 6.3: Verification performance for *minimum* correlation between the training and testing set.

tection of decoherence, as variability is allowed to exist within the burst compared to the previous case. Figure 6.2 illustrates the performance under this scenario. The EER was measured to be *6.3%*.

3. **Minimum window correlation observed within every subject.** This corresponds to the worst case scenario, when no template updating is utilized (or a very low correlation threshold, c_{th}, is used). In this operating instance, the template has low similarity to the biometric signal being measured for potential authentication. Therefore, false rejection is more probable in this case. Figure 6.3 illustrates the respective error and authentication rates for this case. There is a significant increase in the EER, which can be as high as *14.7%*.

It is clear from Figures 6.1 , 6.2 and 6.3 that template updating has no significant impact of FAR, which is random and can only be addressed with appropriate feature extraction. The FRR however drops significantly when a template is updated and matched

against highly correlated inputs, and the same can be observed for EER.

Although there is a significant reduction in FRR when template updating is employed, it is not clear why this is the case. In other words, *why does the ECG signal become decoherent after some time (at the end of a burst), leading to false rejection?* In the next section we apply this algorithm on the affective recordings from the active arousal database, and show that these bursts in the ECG signal correspond to emotional states of the subject, thereby providing a meaningful physical interpretation to what is happening in the updating algorithm.

6.3.2 Biometric Template Updating on Affect Data

The purpose of the following evaluation is to validate that every coherent burst, as detected by the proposed algorithm, describes an true emotional state. Therefore, the subsequent analysis is performed on signals from the active arousal database.

We introduce a measure q_i to describe the system's confidence that burst i represents consistently one emotional state. Every ECG window $x(t)$ in the burst i is labeled as *high* (H) or *low* (L) arousal[1]. Let N_H^i and N_L^i be the number of windows marked as high and low arousal respectively. Then q_i can be calculated as:

$$q_i = \frac{\max(N_H^i, N_L^i)}{N_H^i + N_L^i} \tag{6.2}$$

Essentially, every burst that is identified by the updating algorithm will be shown to have a high q_i, which means that with high confidence each burst corresponds to either high or low arousal. The results are shown in Tables 6.2 and 6.3 which list the q values (in percentage) for all detected bursts of all subjects in the database (for the updating algorithm, moderate thresholds, between 0.8 and 0.85, were selected for every individual). Even though this measure relies on the subjects' self-reports, and discrepancies

[1]This annotation is based on the subject's self report.

are expected, the average state confidence for the bursts is *96.47%*. This performance is illustrative of the accuracy of detecting homogeneous emotional states, which leads to successful template updating.

Returning to our original problem, we are interested in template destabilization for biometric recognition. When a burst is terminated due to decoherence, the biometric template is updated. While the above results show that each burst corresponds to either high arousal or low arousal, the question that naturally arises is the following: *do two consecutive bursts correspond to opposite arousal labels?* In other words, *is change in arousal the only factor that is responsible for burst termination?*

In practice, the template updating algorithm is not only affected by emotional states. A burst may be interrupted due to one of the following three reasons:

1. *A state change* i.e., a transition from one psychological state to another.

2. *Buffer overflow* i.e., when $dsize_{max}$ is reached. $dsize_{max}$ can be adjusted according to the requirements of the application environment. For the present simulation $dsize_{max}$ was set to 10 minutes.

3. *Noise artifacts* (e.g. due to a sudden movement), that are not sufficiently treated by the filter.

Nevertheless, a template update is necessary in all cases. For the purpose of comparison, the verification performance is also estimated over the active arousal dataset for the two cases, namely with and without template updating. Figure 6.4 demonstrates the baseline false acceptance and rejection rates. The equal error rate is *15%*, which is typical for LDA training without template updating.

For every individual in the database, a template is updated every time a new burst is detected. As such, the number of updates is highly dependent on the subject and ranges between 1-8 as listed in Tables 6.2 and 6.3. To quantify the verification accuracy after template updating, false acceptance and rejection rates are estimated for every individual

Subject	**Number of Bursts**	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
1	2	84.6%	100%	-	-	-	-	-	-
2	5	100%	100%	100%	100%	100%	-	-	-
3	4	100%	100%	100%	100%	-	-	-	-
4	2	100%	100%	-	-	-	-	-	-
5	3	83.3%	85.7%	70%	-	-	-	-	-
6	4	100%	100%	100%	100%	-	-	-	-
7	5	100%	89.4%	90.9%	100%	100%	-	-	-
8	3	100%	100%	100%	-	-	-	-	-
9	1	100%	-	-	-	-	-	-	-
10	6	100%	100%	100%	100%	100%	100%	-	-
11	4	100%	100%	100%	100%	-	-	-	-
12	1	100%	-	-		-	-	-	-
13	4	88%	100%	100%	64.3%	-	-	-	-
14	2	100%	100%	-	-	-		-	-
15	3	82.9%	85.7%	100%	-	-	-	-	-
16	1	67.1%	-	-	-	-	-	-	-
17	5	88.8%	70%	88.8%	83.3%	100%	-	-	-
18	2	100%	100%	-	-	-	-	-	-
19	8	100%	76.9%	100%	100%	100%	100%	100%	100%
20	5	100%	100%	100%	100%	100%	-	-	-
21	6	100%	100%	100%	100%	100%	100%	-	-

Table 6.2: State confidence (q) per burst for subjects 1-21.

Subject	**Number of Bursts**	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
22	2	100%	97.2%	-	-	-	-	-	-
23	4	71.4%	100%	100%	100%	-	-	-	-
24	7	100%	100%	100%	100%	100%	100%	100%	-
25	2	95.6%	100%	-	-	-	-	-	-
26	7	62.5%	100%	100%	100%	100%	100%	100%	-
27	5	94.5%	100%	100%	100%	100%	-	-	-
28	5	100%	100%	100%	100%	92.3%	-	-	-
29	1	100%	-	-	-		-	-	-
30	5	100%	100%	100%	100%	100%	-	-	-
31	6	100%	100%	100%	100%	100%	100%	-	-
32	3	100%	100%	100%	-	-	-	-	-
33	7	100%	100%	100%	100%	100%	-	-	-
34	4	100%	100%	100%	100%	-	-	-	-
35	1	90.9%	-	-	-	-	-	-	-
36	4	100%	54.5%	100%	100%	-	-	-	-
37	1	100%	-	-	-	-	-	-	-
38	2	100%	100%	-	-	-	-	-	-
39	2	66.6%	100%	-	-	-	-	-	-
40	4	100%	100%	100%	100%	100%	-	-	-
41	6	60.2%	100%	70%	100%	100%	100%	94.4%	-
42	1	100%	-	-	-	-	-	-	-
43	1	100%	-	-	-	-	-	-	-

Table 6.3: State confidence (q) per burst for subjects 22-43.

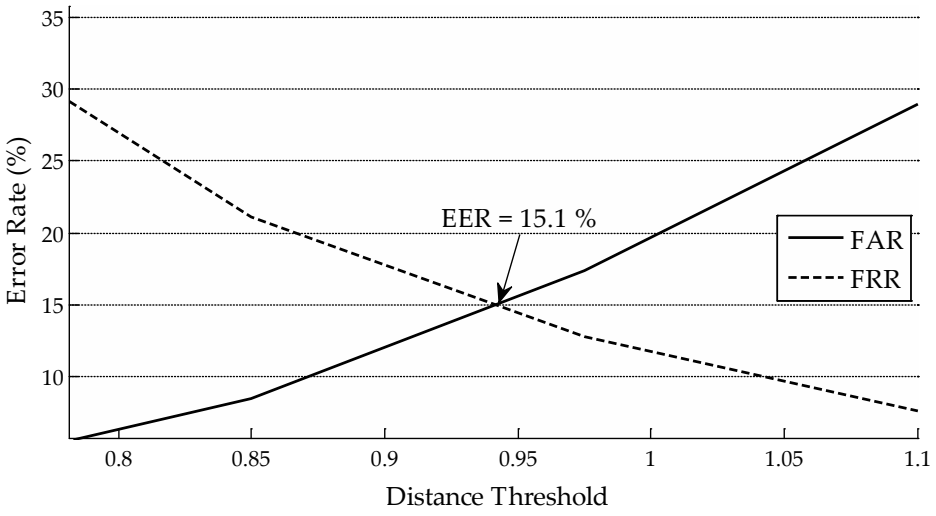


Figure 6.4: Verification performance for the active arousal database, without template updating (EER = 15.1%).

separately. Table 6.4 lists the equal error rates that were achieved with this treatment, for all subjects in the active arousal dataset. The average equal error rate in this case is *3.96%*, which represents a significant improvement from *15%*. Appendix E presents EER results for different coherence thresholds c_{th} and for all individuals in the database.

In addition, Figure 6.5 demonstrates the FA and FR tradeoffs for nine randomly picked individuals. In each simulation false acceptance was computed with comparisons of the updated template of one individual against the remaining subjects in the database.

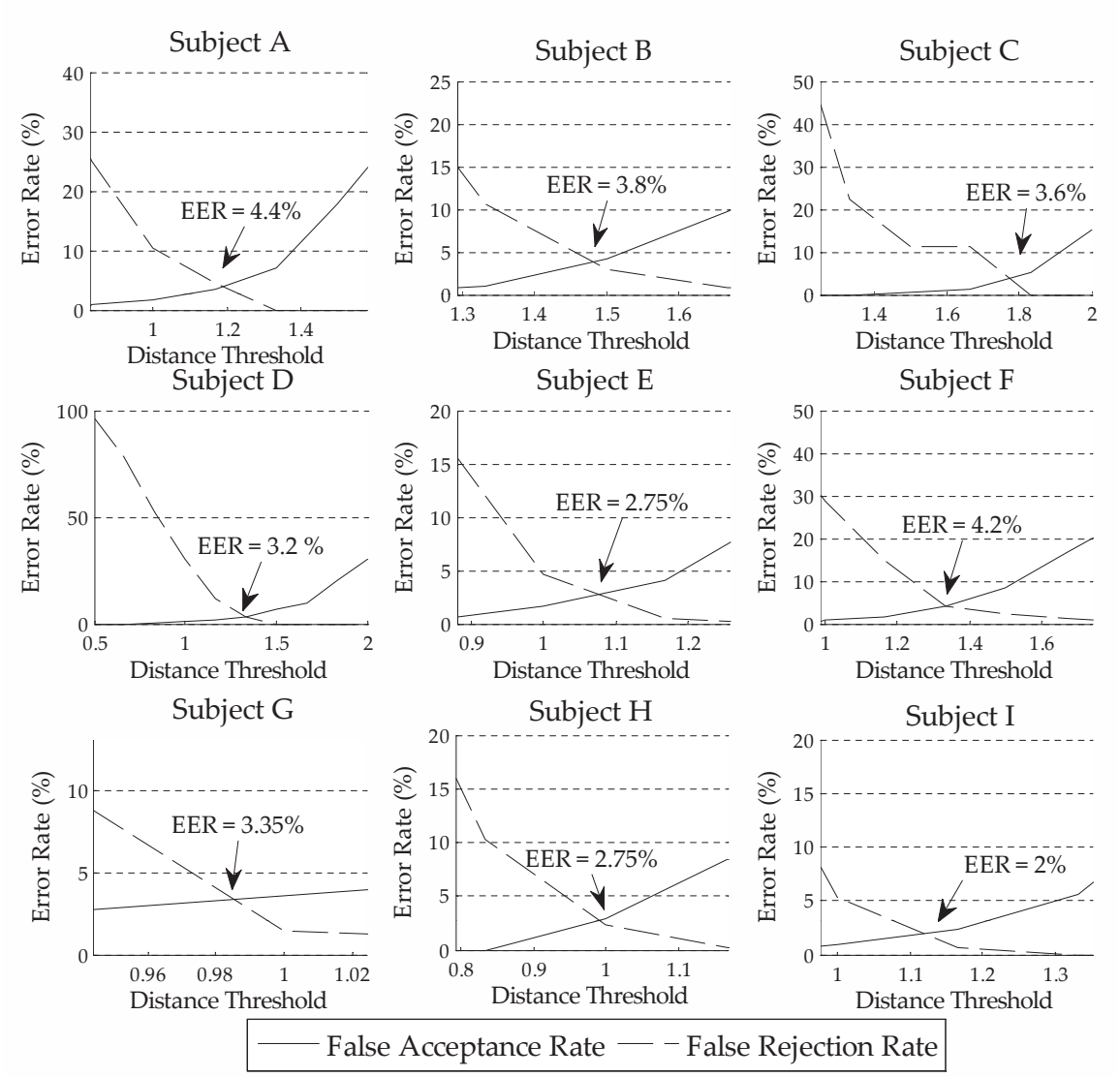


Figure 6.5: Verification performance with template updating for 9 individuals of the active arousal database.

Subject	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**
1	*6%*	11	*4.3%*	21	*2.65%*	31	*5%*	41	*0.2%*
2	*4.2%*	12	*3.6%*	22	*3%*	32	*5.16%*	42	*0%*
3	*8%*	13	*1.1%*	23	*5.65%*	33	*2.65%*	43	*1.35%*
4	*4.4%*	14	*4.3%*	24	*3.95%*	34	*9.2%*		
5	*4%*	15	*3.2%*	25	*3.38%*	35	*8.9%*		
6	*0%*	16	*2.5%*	26	*4.04%*	36	*1.85%*		
7	*0.5%*	17	*2.75%*	27	*3.35%*	37	*7.65%*		
8	*5.5%*	18	*4.63%*	28	*1.46%*	38	*2%*		
9	*0.05%*	19	*4.2%*	29	*2.75%*	39	*2.95%*		
10	*3.8%*	20	*2.3%*	30	*0.5%*	40	*13.34%*		

Table 6.4: Equal error rate for each individual in the active arousal database, after template updating. Mean equal error rate is 3.96%

6.4 Chapter Summary

This chapter presented a solution to address the problem of template destabilization in ECG biometrics. The perils of ignoring the destabilization of the template were first demonstrated over ECG recordings from the long-term database. Specifically, it was shown that if no special consideration is paid to this problem, the equal error rate can be as high as *14.7%*.

In monitoring environments where the ECG signal is constantly acquired, it is critical to address this problem. In most real life settings, the emotional activity that affects the ECG signal is characterized by high arousal, since the individual is usually actively engaged in some everyday task. Moreover, Chapter 4 demonstrated the feasibility of detecting human emotion from the ECG in such regimes, which tells us that there is a significant risk of template desabilization. Automatic detection of template destabilization due to psychological variations is therefore very important in order to control false

rejection.

The template updating methodology that was presented in this chapter was also evaluated over the active arousal database of ECG signals. The purpose of this analysis was two-fold. First, it was shown that the accumulated bursts that were detected with the proposed solution correspond to homogeneous emotional states with high probability (see Tables 6.2) and 6.3). Furthermore, the verification accuracy was evaluated for every subject when the respective biometric template was updated in the beginning of a burst. On average an equal error rate of *3.96%* was achieved, which represents a dramatic reduction from the EER of *15%* for the particular dataset in the absence of template updating.

Chapter 7

Conclusion

7.1 Thesis Summary

In this thesis, we address the problem of ECG-based human recognition from a theoretical, algorithmic and application point of view.

The core idea of the proposed solutions is the use of the autocorrelation of the ECG signal in conjunction with Linear Discriminant Analysis (AC/LDA). The AC/LDA algorithm was enhanced to incorporate the periodicity transform as a signal quality assessment tool. We demonstrated that the enhanced algorithm decreases the intra-subject variability by *25%* for a dataset of short-term ECG recordings from 52 healthy volunteers.

The enhanced AC/LDA algorithm was subsequently tested in three different scenarios, each of which envisions a unique application environment. The scenarios considered are A) small-scale recognition, B) large-scale recognition and C) recognition in a distributed system. Scenario B presents the most challenging application setting because in this environment the population of enrollees is unknown at the time of LDA training. On the other hand, in scenario A the application environment is controlled in the sense that the LDA algorithm is trained on the exact enrollee recordings. Finally, scenario C presents

the opportunity of performing the biometric recognition locally on a smart device, with the advantage of personalizing the system to the intra-subject variability that is expected from a particular user.

Scenario B presents a challenge with regard to the machine-learning algorithm, namely the absence of an appropriate training set. We address this problem by designing a generic pool of ECG signals to train the LDA. The underlying idea is that the system can benefit from learning ECG morphologies from the general population, which is essentially a way to control false rejection. However, false acceptance continues to remain a significant problem for this scenario because the LDA has not been trained on morphologies of the particular enrollees. The equal error rare for this scenario is *45.5%*, which discourages deployment of ECG biometrics in large-scale recognition environments.

A better performance is achieved for Scenarios A and C. Since the population of enrollees is known at the time of LDA training, the equal error rate drops to *12%* for Scenario A and *10%* (on average) for Scenario C. These results encourage the use of the ECG signal in human recognition, and also emphasize the need for machine learning as a way to address the intra-subject variability of this biometric modality.

This thesis also examines the ECG signal from a psychological point of view. The rationale is that in the absence of noise or physical activity, an ECG template may still destabilize with time due to variations in the emotional state. This issue becomes critically important in monitoring environments wherein the signal is used to continuously authenticate the user's identity.

An important contribution of this thesis is the identification of user emotional states which present risks to biometric recognition. This problem is addressed from an affective computing point of view. Appropriate experimental setups were designed to collect ECG signals under particular emotional states. The experiments included inspection of visual stimuli (passive arousal) and video games (active arousal). The signal is examined under passive and active arousal as well as under positive and negative valence.

The empirical mode decomposition (EMD) was employed for emotional pattern classification. The algorithm was first adapted to the ECG signal using the bivariate extension on a real and a synthetic ECG segment. Classification was then carried out using features that picture the local frequency and oscillation of the intrinsic mode functions. We demonstrated that active arousal may be detected better than passive arousal, with an overall classification rate of *78.43%*. This finding is critical for real life deployment of ECG biometrics wherein active arousal is anticipated to be a dominant factor.

Finally, this thesis addressed the problem of template destabilization due to psychological variations. The proposed solution accumulates ECG segments into bursts until a decoherence is detected with respect to the previously designed biometric template. Once the coherence of the accumulated duration is disturbed a new biometric template is created. The performance was evaluated over ECG signals from the active arousal database. There were two significant observations. It was observed that with high probability (*96.47%*), the detected bursts correspond to distinct emotional states. Secondly, that the equal error rate decreases to *3.96%* if the biometric template is updated in the beginning of the corresponding burst.

The findings in this thesis make a strong case for the deployment of ECG biometrics in human recognition. Although the signal has been under biometric investigation only within the last decade, the state of the art performance implies that this modality will soon find its niche in the biometrics world.

7.2 Future Work

There are a number of areas touched upon in this thesis that provide direction for future research.

7.2.1 Online State Detection and Prediction.

This thesis proposed a solution to automatic template updating, by examining the coherence of accumulated segments of an ECG signal. This is the most basic approach to state detection, which essentially corresponds to offline change estimation. Another direction to explore is the use of more sophisticated change detectors which operate on the signal in an online fashion. Likelihood ratio tests can be considered using the *Generalized Likelihood Ratio* or the *Marginalized Likelihood Ratio*.

In addition, environmental cues and/or additional modalities can be incorporated to the system and help predict an emotional change on the ECG signal. In controlled environments, for example in an online teleconference system, one can analyze vocal cues (or the semantics of an on-going discussion) to estimate an *a priori* emotional probability which can then been sought in the ECG signal.

7.2.2 Investigation of Acceptable Waiting Periods

In this thesis, the AC/LDA recognition algorithm operated on ECG segments of finite length. The underlying idea is that real life systems require identity assessments as fast as possible. Therefore, the fundamental waiting time for the proposed solution was set to 5 seconds in order to accommodate the requirements of both the feature extractor (the autocorrelation,) and the user convenience with respect to waiting times.

However, fixing the fundamental waiting time is not the only option to address ECG biometric recognition. As the signal flows continuously into the system, and the amount od identifying data increases it is anticipated that the probability of positive recognition will increase as well. In fact, there are environments where immediate subject recognition is not critical (for instance, patient monitoring within the hospital), and a possibility is therefore presented for higher recognition accuracy.

In this regard, one future direction is to define measures that express the system's

confidence in providing an accurate identity decision, based on the variability of accumulated ECG segments. Practically, this requires a feature or decision level fusion of the information as it becomes available to the system. It is for instance anticipated, that voting among accumulated segments that form a burst will significantly reduce false rejection rates.

7.2.3 Addressing Privacy Concerns.

With the potential of using the ECG signal for human recognition the risk of privacy intrusion becomes prominent. Traditionally, only designated health care practitioners have access to this signal, as it may reveal current and past medical conditions. Once this information is compromised, the results can be catastrophic for people's integrity.

Furthermore, as this thesis also demonstrated, the ECG signal can provide hints of psychological activity. This increases the privacy risk as the signal can directly provide information on the instantaneous emotional activity of the individual.

Although, other biometric modalities also face the same problem, the risk that ECG encompasses may hold it from being accepted by the public as a safe and secure means of authentication. It is important therefore to examine methodologies that address this problem from the users' point of view. Biometric encryption is one solution to this problem, wherein the biometric modality is encrypted with keys that are provided by the biometric itself and can only be unlocked once the same biometric is presented again. However, the time-dependent nature of the ECG presents a great challenge to biometric encryption on this signal, as the feature space needs to be first optimized for this task.

7.2.4 Fusion of Medical Biometrics

The solutions that are presented in this thesis can be easily extended to other medical signals such as the blood pressure, the electroencephalogram, the phonocardiogram and the photoplethysmogram. Although the ECG signal is the most widely explored medical

biometric, the above biosignals may also be subject specific.

In principal, using more than one modalities for human recognition increases the probability of finding the right math in a gallery set. The fact that medical biometrics have the same intrinsic properties (liveness detection, universality, permanence) motivates their fusion into a multi-modal biometric system. Once every signal is established as a biometric modality individually, the next step will be the design of fusion mechanisms that can strategically combine the discriminative information from each of them. Fusion can be performed at three different levels i.e., at a raw datal level, at a feature level or at a decision level.

Appendix A

The Empirical Mode Decomposition

The Empirical Mode Decomposition algorithm has been proposed by Huang *et al.* [90]. It is a non-linear technique to decompose a non-stationary signal, after identifying its oscillatory modes. There are three underlying assumptions:

1. The analyzed signal should have as many local extrema as possible.
2. The time scale is defined as the time passing between two extrema.
3. If the signal does not have apparent extrema, then differentiation is applied to reveal them.

For EMD, oscillatory modes correspond to *Intrinsic Mode Functions* (IMF). The IMFs represent the various intrinsic time scales of the signal and satisfy two conditions:

1. Their number of extrema and the zero crossings must be equal or doffer at most by one.
2. The average value of the envelope defined by the local maxima and local minima is zero at any time point.

As explained in [90], most practically encountered signals do not satisfy the properties of an IMF, but can be expanded into a finite set of IMFs.

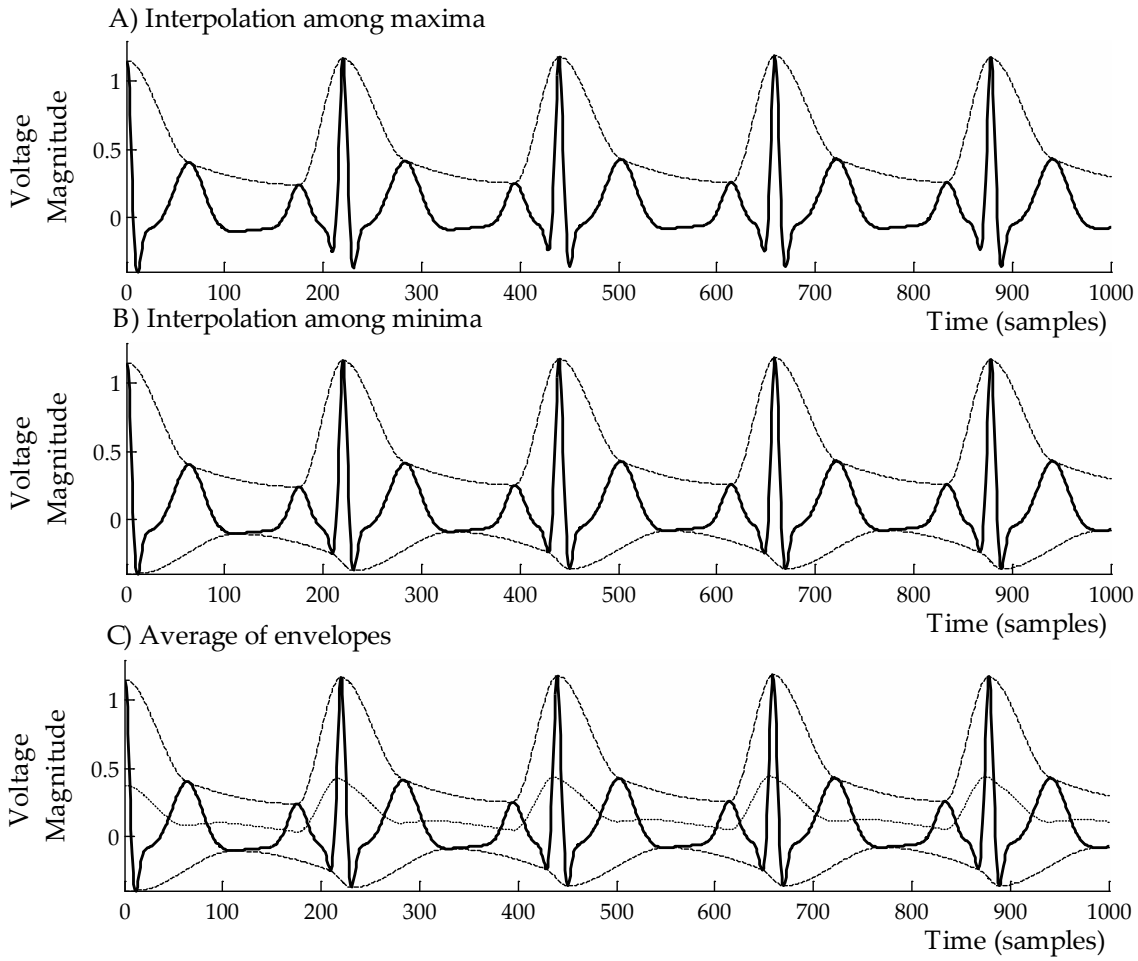


Figure A.1: EMD steps on an ECG signal.

With EMD the a signal $x(t)$ can be expressed as:

$$x(t) = \sum_{i=1}^{M} c_i(t) + r(t) \tag{A.1}$$

where $c_i(t)$ for $i = 1, ...M$ is the set of IMFs and $r(t)$ is the final residual. The number of IMFs can not be predefined and it depends on the volume of the oscillatory activity in x_t).

More specific, given a signal $x(t)$ the EMD algorithm operates as follows:

1. Detect local maxima $x_{max}(i)$ and minima $x_{min}(j)$ of $x(t)$.

2. Interpolate among $x_{max}(i)$ to get an upper envelope $x_{up}(t)$, and $x_{low}(t)$ for minima respectively.

3. Compute the average of envelopes $m(t) = \frac{x_{up}(t)+x_{low}(t)}{2}$.

4. Subtract from signal $u(t) = x(t) - m(t)$.

5. Iterate for $x(t) = u(t)$.

An example of this procedure on the ECG signal is depicted in Figure A.1. EMD describes essentially a *sifting* process, which is terminated when $u(t)$ meets the IMF criteria. If it does, $u(t)$ will be describing a underlying oscillation of $x(t)$, referred to herein as $d(t)$. EMD continues with sifting on the residual $r(t) = x(t) - d(t)$. The original signal can then be expressed as:

$$x(t) = \sum_{i=1}^{N-1} d_i(t) + r(t) \tag{A.2}$$

where $d_i(t)$ denotes the i^{th} IMF extracted from the signal $x(t)$ and $r(t)$ is the final residual. Note that by definition, $r(t)$ is not an IMF.

Higher order IMFs describe slower oscillations, and are thus dominated by the smaller frequencies of $x(t)$. An example of EMD decomposition for an ECG segment is depicted in Figure A.2.

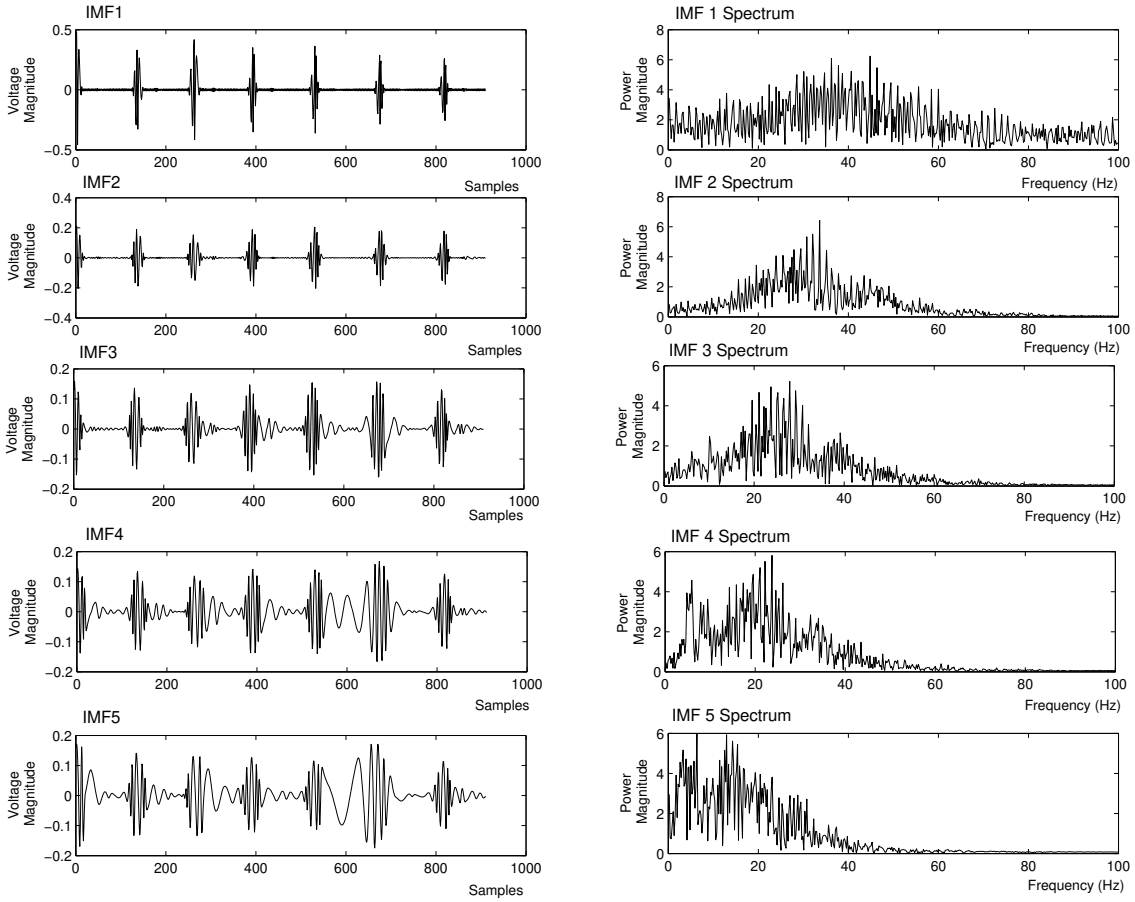


Figure A.2: EMD analysis for a real ECG segment (5 out of 16 IMFs).

Appendix B

Human Emotion Models

Research on emotional intelligence is inevitably linked with studying emotional organization and classification. Psychologists have long been trying to provide a rational description for emotional clusters, if any such exists. In principal two scientific models have been suggested to describe emotions namely the *Affective Dimensional Grid* (ADM) and the *Discrete Emotional Models* (DEM).

The discrete emotional models suggest that there exists a set of distinct emotions, which are clearly specified and separated. Ekman [58] suggested that emotions are only those psychological states which conform to specific criteria. Examples of these requirements, are for the affective state to be universal among people, to have a distinctive physiological expression with a quick onset and brief duration and other. Standard emotions considered usually are happiness, surprise, anger, dislike, fear,and sadness.

The first dimensional model for the description of emotions was proposed by Russel *et a.* [5] in 1989. The affect grid idea suggests that emotions are not discrete but continuous, and can be projected in a two dimensional space. The first dimension pictures the *valence* of the emotion and ranges between pleasure and displeasure. The second dimension described the *valence* and ranges between relaxation and high arousal. Figure B.1 illustrates the 2D affect space (AV) and shows the approximate coordinates for some

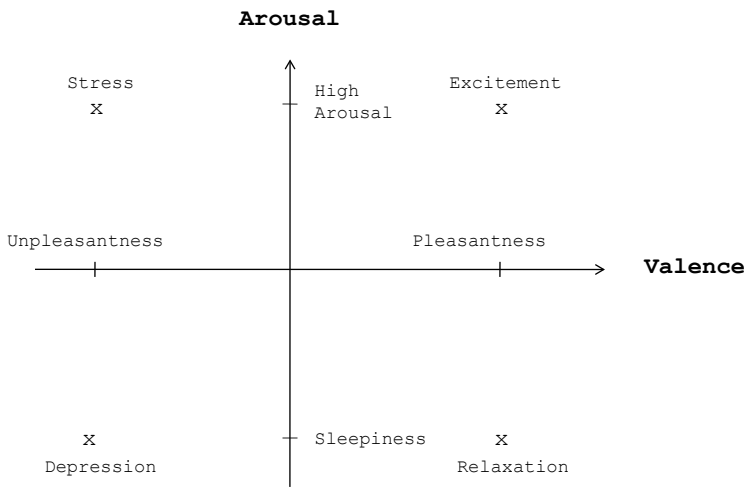


Figure B.1: The Arousal-Valence (AV) plane. Reproduced from [5].

well established emotional states. Lang [100], categorized real life pictures with respect to the elicited arousal and valence, and managed to create a mapping between the pictures and the AV plane.

Both theories have advantages and disadvantages. The explicit specification of emotions might be too difficult as they are expressed in various ways even by the same person under different environments. On the other hand, the AV plane assumes independence between arousal and valence which is practically uncertain. However, further analysis of the psychological aspect of emotions is beyond the scope of this proposal.

Appendix C

Affective Computing Features in Prior Art

Reference Signal	**Feature**
GSR [a]	Detrended time series (subtraction of a time varying sample mean)
GSR	Unbiased sample variance of the detrended signal
BVP [b]	Pinch: Difference between upper and lower envelope
BVP	Peak to peak intervals (related to heart rate)
BVP	Local variance of coefficients of the 3-level wavelet expansion

Table C.1: Features for classification used by Scheirer *et al.* in [2].

[a] Galvanic Skin Response
[b] Blood Volume Pressure

Reference Signal	**Estimator**	**Description**
Heart rate H	$f_1 = \frac{1}{N}\sum_{n=1}^{N} b_n$	b is acquired by convolving with a Hanning window on H. This handles the fluctuations on H by making the heart rate waveform smoother.
Heart rate H	$f_2 = \frac{1}{N-1}\sum_{n=1}^{N-1}(b_{n+1} - b_n)$	This measure describes the average acceleration or deceleration of the heart rate.
Skin Conductivity S	$f_3 = \frac{s\min(s)}{\max(s)-\min(s)}$	s is the result of the convolution between S and a Hanning window. This eliminates high frequency noise effects in S. The maximum and minimum values are with respect to the whole day's observations.
Skin Conductivity S	$f_4 = \frac{1}{N-1}\sum_{n=1}^{N-1}(s_{n+1} - s_n)$	A statistical measure which refers to the mean of the first differences of the smoothed S.
Respiration Rate R	$f_5 = \frac{1}{N}\sum_{n=1}^{N} r_n$	Where $r = R - \frac{1}{N_{day}}\sum_{n=1}^{N_{day}} R_n$. This measure accounts for variations due to tightness of the sensor placement on different days.
Respiration Rate R	$f_6 = \frac{1}{N-1}\sum_{n=1}^{N}(r_n - \frac{1}{N_{day}}\sum_{n=1}^{N_{day}} R_n)^2$	A statistical measure of r.
Respiration Rate R	f_7	Power spectral density amplitude of R in the interval 0.0-0.1 Hz.
Respiration Rate R	f_8	Power spectral density amplitude of R in the interval 0.1-0.2 Hz.
Respiration Rate R	f_9	Power spectral density amplitude of R in the interval 0.2-0.3 Hz.
Respiration Rate R	f_{10}	Power spectral density amplitude of R in the interval 0.3-0.4 Hz.

Table C.2: Typical features for classification used by Picard *et al.* in [1].

Feature No	**Reference Signal**	**Feature**
1	EMG [a]	Normalized mean
2	R [b]	Normalized mean
3	R	Normalized variance
4	HR [c]	Normalized mean
5	HR [d]	Normalized variance
6	GSR foot [e]	Normalized mean
7	GSR foot	Normalized variance
8	GSR hand	Normalized mean
9	GSR hand	Normalized variance
10	R	PSD in 0-0.1Hz
11	R	PSD in 0.1-0.2Hz
12	R	PSD in 0.2-0.3Hz
13	R	PSD in 0.3-0.4Hz
14	GSR foot	Sum or *Orienting Response* [f] magnitudes
15	GSR foot	Sum of Orienting Response durations
16	GSR foot	Sum of areas under Orienting Responses
17	GSR foot	Number of Orienting Responses
18	GSR hand	Sum or *Orienting Response* magnitudes
19	GSR hand	Sum of Orienting Response durations
20	GSR hand	Sum of areas under Orienting Responses
21	GSR hand	Number of Orienting Responses
22	HRV [g]	Ratio of low to high spectral energies

Table C.3: Features for *Analysis I* classification used by Healey *et al.* in [3].

[a] Electromyogram
[b] Respiration
[c] Heart Rate
[d] Heart Rate
[e] Galvanic Skin Response
[f] An Orienting response is a sudden, high frequency pick in GSR.
[g] Heart Rate Variability

Appendix D

Heart Rate Variability

The Heart Rate Variability (HRV) is a measure that describes the heart-beat occurrence times on the ECG signal. Due to the simplicity in the collection and estimation, the HRV is used widely in order to assess the functions of the autonomic nervous system and related diseases. The inter-beat intervals that are analyzed for HRV estimation are not necessarily acquired from the ECG signal, although this is the most accurate way to perform this analysis.

The first step to HRV estimation is the localization of the R peaks, as shown in Figure D.1. Then, the intervals between successive R peaks are estimated as depicted in Figure D.2, and the HRV can be expressed as either the instantaneous change of this time-series or simply the standard deviation. Traditionally, the P peaks were preferred for this analysis because their occurrence is directly linked to the SA node, however the detection of a P wave may be inaccurate due to its small amplitude. Since the RR interval is relatively stable, the R waves are widely used for HRV estimation. Most of the HRV analyses take place in frequency domain. Typically, the spectrum of the HRV is divided in two parts as follows:

1. *Low Frequency*: 0.04Hz - 0.15Hz

2. *High Frequency*: 0.15Hz - 0.4Hz

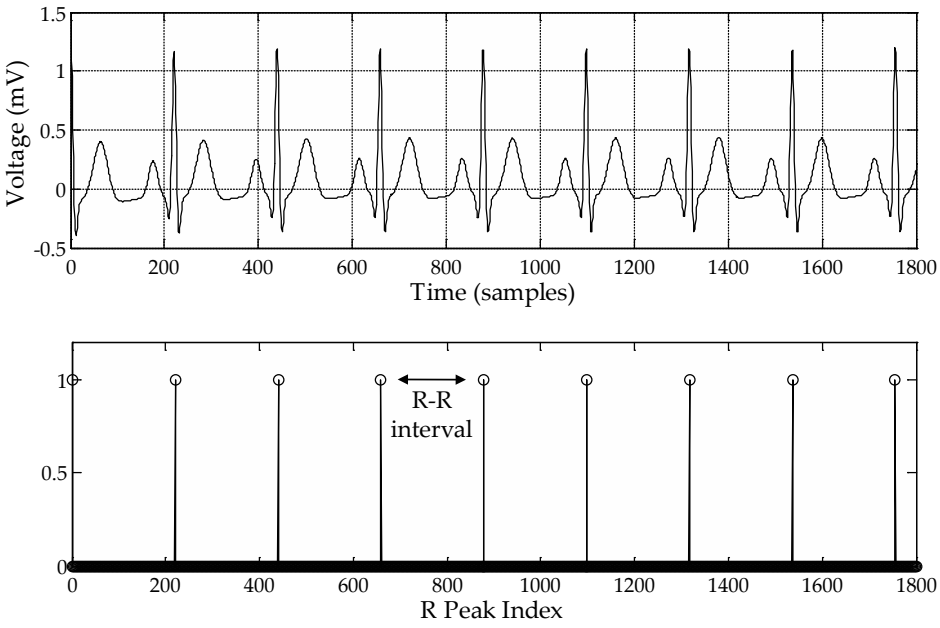


Figure D.1: R peak detection for the HRV estimation.

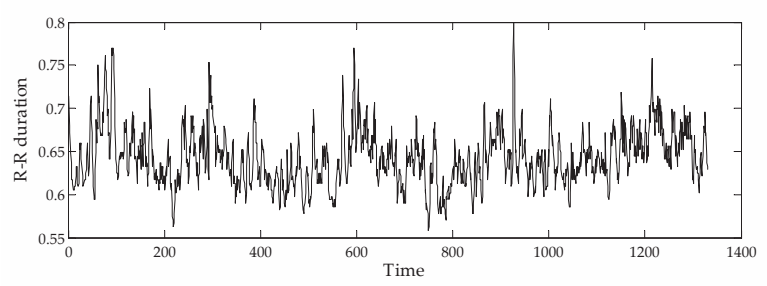


Figure D.2: *R-R* time series used for HRV estimation.

The power measured in these two intervals is highly correlated with processes of the autonomic nervous system. Sympathetic activity is associated with the low frequency band, and parasympathetic with the high frequency. The ratio $\frac{LF}{HF}$ between the two spectral power measures is an indication of autonomic balance.

Given the relation of the HRV with the autonomic nervous system, this signal has been associated with a number of conditions, and has been studied in different fields:

1. In affective computing, the HRV is directly associated with the degree of engagement in a task. As a measure of mental effort, it is used in high-stress environments

(for example air-traffic control) for assessing the mental state of the users [101, 102]. In this field, the heart rate has been also shown to differentiate between positive and negative emotions [101]. Typically, the spectral characteristics of the HRV series are used in affective computing. Instead of just two frequency bands, the spectrum is divided in parts i.e., *Very low frequencies* (VLF: 0.003Hz - 0.4Hz), *low frequencies* (LF: 0.04Hz - 0.15Hz)and *High frequencies* (HF: 0.15Hz - 0.4Hz). From these subband spectra, the dominant powers and frequencies are estimated for each band, by integrating the power spectral densities. The ratios of the dominant frequencies are used to distinguish sympathetic from parasympathetic effects [101].

2. The HRV has been shown to decrease in patients with acute myocardial infraction [103].

3. Disorders of the central and peripheral nervous systems have effects on the HRV. The spectral characteristics of the HRV provide insight on the neurological condition of patients and can help diagnose brain damage or depression [103].

4. Using the HRV, it was shown that smokers have increased sympathetic activity [103].

However, the information that the HRV provides is limited to the heart-beat occurrence, while no information on the morphology of the ECG is conveyed.

Appendix E

Template Update Results

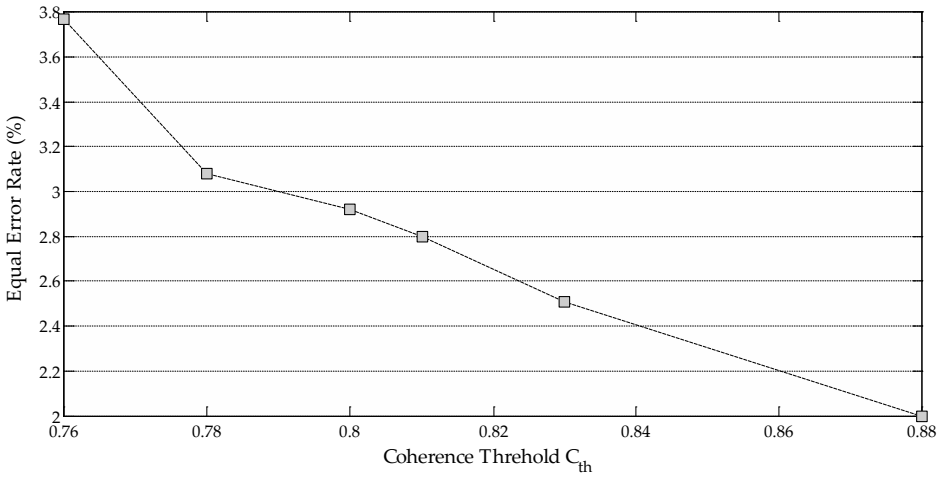


Figure E.1: Equal error rates of the template updating algorithm for various coherence thresholds c_{th}. As the threshold increases, better coherence is imposed on the bursts which leads to more frequent template updating and subsequently smaller EER.

Subject	EER	Subject	EER	Subject	EER	Subject	EER	Subject	EER
1	*7.24%*	11	*3.36%*	21	*2.06%*	31	*0.99%*	41	*0.57%*
2	*3.74%*	12	*4.04%*	22	*1.12%*	32	*3.75%*	42	*0.2%*
3	*6.13%*	13	*0.56%*	23	*5.02%*	33	*1.35%*	43	*8.08%*
4	*5.6%*	14	*3.52%*	24	*4.01%*	34	*7.05%*		
5	*3.69%*	15	*2.47%*	25	*2.93%*	35	*8.89%*		
6	*0.069%*	16	*1.84%*	26	*5.72%*	36	*7.44%*		
7	*1.63%*	17	*2.5%*	27	*4.77%*	37	*6.92%*		
8	*4.3%*	18	*4.48%*	28	*0.96%*	38	*1.82%*		
9	*3.46%*	19	*4.49%*	29	*2.92%*	39	*5.97%*		
10	*3.48%*	20	*1.96%*	30	*0.21%*	40	1*10.96%*		

Table E.1: Equal error rate for $c_{th} = 0.76$. Mean equal error rate is 3.77%

Subject	EER	Subject	EER	Subject	EER	Subject	EER	Subject	EER
1	*6.76%*	11	*3.14%*	21	*0.27%*	31	*0.54%*	41	*0.19%*
2	*3.76%*	12	*4.04%*	22	*1.12%*	32	*2.83%*	42	*0.174%*
3	*8.87%*	13	*0.57%*	23	*5.42%*	33	*1.01%*	43	*2.49%*
4	*4.14%*	14	*3.19%*	24	*0.51%*	34	*7.71%*		
5	*3.81%*	15	*2.47%*	25	*3.05%*	35	*8.89%*		
6	*0.08%*	16	*1.84%*	26	*5.71%*	36	*2.24%*		
7	*1.63%*	17	*2.1%*	27	*3.92%*	37	*6.92%*		
8	*4.3%*	18	*2.94%*	28	*0.96%*	38	*1.07%*		
9	*3.47%*	19	*4.55%*	29	*2.92%*	39	*5.97%*		
10	*2.17%*	20	*2.28%*	30	*0.21%*	40	1*2.38%*		

Table E.2: Equal error rate for $c_{th} = 0.78$. Mean equal error rate is 3.08%

Subject	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**
1	*9.31%*	11	*3.25%*	21	*0.46%*	31	*0.59%*	41	*0.19%*
2	*3.73%*	12	*0.1%*	22	*1.11%*	32	*23.42%*	42	*0.174%*
3	*8.81%*	13	*0.57%*	23	*5.42%*	33	*1.42%*	43	*3.66%*
4	*4.14%*	14	*3.19%*	24	*0.98%*	34	*7.71%*		
5	*3.81%*	15	*2.47%*	25	*2.86%*	35	*8.89%*		
6	*0.08%*	16	*1.84%*	26	*4.57%*	36	*1.98%*		
7	*1.21%*	17	*2.11%*	27	*3.73%*	37	*6.92%*		
8	*4.3%*	18	*2.72%*	28	*0.96%*	38	*1.16%*		
9	*0.87%*	19	*4.51%*	29	*2.92%*	39	*5.97%*		
10	*2.27%*	20	*2.94%*	30	*0.09%*	40	*12.38%*		

Table E.3: Equal error rate for $c_{th} = 0.8$. Mean equal error rate is 2.92%

Subject	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**
1	*9.31%*	11	*3.01%*	21	*0.46%*	31	*0.84%*	41	*0.19%*
2	*3.73%*	12	*0.1%*	22	*0.89%*	32	*3.87%*	42	*0.17%*
3	*8.81%*	13	*0.95%*	23	*5.93%*	33	*1.26%*	43	*2.84%*
4	*4.14%*	14	*2.2%*	24	*0.85%*	34	*7.71%*		
5	*3.81%*	15	*3.92%*	25	*2.86%*	35	*8.89%*		
6	*0.08%*	16	*1.84%*	26	*3.79%*	36	*1.98%*		
7	*1.21%*	17	*2.12%*	27	*3.29%*	37	*4.71%*		
8	*4.3%*	18	*2.72%*	28	*0.38%*	38	*1.13%*		
9	*0.03%*	19	*4.72%*	29	*1.46%*	39	*2.72%*		
10	*2.35%*	20	*2.34%*	30	*0.1%*	40	*12.87%*		

Table E.4: Equal error rate for $c_{th} = 0.81$. Mean equal error rate is 2.8%

Subject	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**
1	*9.05%*	11	*2.65%*	21	*0.64%*	31	*0.88%*	41	*0.11%*
2	*2.96%*	12	*0.1%*	22	*0.89%*	32	*6.07%*	42	*0.17%*
3	*0.1%*	13	*0.95%*	23	*5.93%*	33	*1.25%*	43	*2.84%*
4	*4.45%*	14	*1.45%*	24	*0.61%*	34	*6.76%*		
5	*3.81%*	15	*2.35%*	25	*2.86%*	35	*8.89%*		
6	*0.08%*	16	*1.84%*	26	*3.65%*	36	*1.98%*		
7	*0.95%*	17	*2.04%*	27	*1.7%*	37	*4.71%*		
8	*2.98%*	18	*1.59%*	28	*0.38%*	38	*1.13%*		
9	*0.03%*	19	*5.18%*	29	*1.46%*	39	*7.41%*		
10	*2.22%*	20	*2.34%*	30	*0.1%*	40	1*2.87%*		

Table E.5: Equal error rate for $c_{th} = 0.83$. Mean equal error rate is 2.51%

Subject	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**	**Subject**	**EER**
1	*3.83%*	11	*2.28%*	21	*0.05%*	31	*0.88%*	41	*0.61%*
2	*3.59%*	12	*0.01%*	22	*1.12%*	32	*3.46%*	42	*0.17%*
3	*0.1%*	13	*0.93%*	23	*1.47%*	33	*1.84%*	43	*1.25%*
4	*3.55%*	14	*1.68%*	24	*0.91%*	34	*6.76%*		
5	*3.4%*	15	*2.19%*	25	*2.86%*	35	*8.89%*		
6	*0.06%*	16	*1.84%*	26	*3.61%*	36	*1.76%*		
7	*0.67%*	17	*2.26%*	27	*1.7%*	37	*3.94%*		
8	*2.98%*	18	*1.59%*	28	*0.31%*	38	*1.78%*		
9	*0.01%*	19	*4.83%*	29	*1.68%*	39	*0.01%*		
10	*1.86%*	20	*2.7%*	30	*0.01%*	40	1*2.63%*		

Table E.6: Equal error rate for $c_{th} = 0.88$. Mean equal error rate is 2%

Bibliography

[1] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[2] J. Scheirer, R. Fernandez, J. Klein, and R. Picard, "Frustrating the user on purpose: a step toward building an affective computer," *Interacting with Computers*, vol. 14, no. 2, pp. 93–118, Feb. 2002.

[3] J. Healey and R. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 6, no. 2, pp. 156–166, June 2005.

[4] "The PTB diagnostic ECG database, national metrology institute of germany," http://www.physionet.org/physiobank/database/ptbdb/.

[5] J. Russell, A. Weiss, and G. Mendelsohn, "Affect grid: A single-item scale of pleasure and arousal," *Journal of Personality and Social Psychology*, vol. 57, no. 3, pp. 493–502, Sept. 1989.

[6] "The US Federal Commission report, February 2010," http://www.ftc.gov/sentinel/reports/sentinel-annual-reports/sentinel-cy2009.pdf.

[7] "Peter allen, calais migrants mutilate fingerprints to hide true identity, daily mail," http://www.dailymail.co.uk/news/worldnews/article-1201126/Calais-migrants- mutilate-fingertips-hide-true-identity.html.

[8] F. Agrafioti and D. Hatzinakos, "ECG biometric analysis in cardiac irregularity conditions," *Signal, Image and Video Processing*, pp. 1863–1703, 2008.

[9] ——, "ECG based recognition using second order statistics," in *Proceedings of 6th Annual Communication Networks and Services Research Conference*, Halifax, May. 2008, pp. 82 –87.

[10] L. Sornmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Elsevier, 2005.

[11] L. Green, R. Lux, C. H. R. Williams, S. Hunt, and M. Burgess, "Effects of age, sex, and body habitus on QRS and ST-T potential maps of 1100 normal subjects," *Circulation*, vol. 85, pp. 244–253, 1985.

[12] H. Draper, C. Peffer, F. Stallmann, D. Littmann, and H. Pipberger, "The corrected orthogonal electrocardiogram and vectorcardiogram in 510 normal men (Frank lead system)," *Circulation*, vol. 30, pp. 853–864, 1964.

[13] T. Pilkington, R. Barr, and C. L. Rogers, "Effect of conductivity interfaces in electrocardiography," *Springer New York.*, vol. 30, pp. 637–643, 2006.

[14] H. Larkin and S. Hunyor, "Precordial voltage variation in the normal electrocardiogram," *J. Electrocardiology*, vol. 13, pp. 347–352, 1980.

[15] G. Kozmann, R. Lux, and L. Green, "Sources of variability in normal body surface potential maps," *Circulation*, vol. 17, pp. 1077–1083, 1989.

[16] ——, "Geometrical factors affecting the interindividual variability of the ECG and the VCG," *J. Electrocardiology*, vol. 33, pp. 219–227, 2000.

[17] R. Hoekema, G.Uijen, and A. van Oosterom, "Geometrical aspect of the interindividual variaility of multilead ECG recordings," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 551–559, 2001.

[18] B. P. Simon and C. Eswaran, "An ECG classifier designed using modified decision based neural network," *Comput. Biomed. Res.*, vol. 30, pp. 257–272, 1997.

[19] J. T. Catalano, *Guide to ECG analysis*. Philadelphia: J.B. Lippincott, 1993.

[20] R. Shouldice, C. Heneghan, P. Nolan, and P. Nolan, "PR and PP ECG intervals as indicators of autonomic nervous innervation of the cardiac sinoatrial and atrioventricular nodes," in *Proceedings of 1st Int. Conference on Neural Eng.*, March 2003, pp. 261 – 264.

[21] R. Virtanen, A. Jula, J. Salminen, L. Voipio-Pulkki, H. Helenius, T. Kuusela, and J. Airaksinen, "Anxiety and hostility are associated with reduced baroreflex sensitivity and increased beat-to-beat blood pressure variability," *Psychosomatic Medicine*, vol. 65, pp. 751 – 756, 2003.

[22] S. Booth-Kewley and H. S. Friedman, "Psychological predictors of heart disease: A quantitative review," *Psychological Bulletin*, vol. 101, no. 3, pp. 343 – 362, 1987.

[23] G. H. E. G. M. Richter, A. Friedrich, "Task difficulty effects on cardiac activity," *Psychophysiology*, no. 45, pp. 869–875, 2008.

[24] J. Blascovich, M. Seery, C. Mugridge, R. Norris, and M. Weisbuch, "Predicting athletic performance from cardiovascular indexes of challenge and threat," *Journal of Experimental Social Psychology*, vol. 40, no. 5, pp. 683 – 688, 2004.

[25] H. Ue, I. Masuda, Y. Yoshitake, Y. Inazumi, and T. Moritani, "Assessment of cardiac autonomic nervous activities by means of ECG R-R interval power spectral

analysis and cardiac depolarization-repolarization process," *Annals of Noninvasive Electrocardiology*, vol. 5, no. 4, pp. 336–345, 2000.

[26] W. Wasmund, E. Westerholm, D. Watenpaugh, S. Wasmund, and M. Smith, "Interactive effects of mental and physical stress on cardiovascular control," *J. Appl. Physiol*, vol. 92, pp. 1828–1834, 2002.

[27] R. Pramila, J. Sims, R. Brackin, and N. Sarkar, "Online stress detection using psychophysiological signals for implicit human-robot cooperation," *Robotica*, vol. 20, no. 6, pp. 673–685, 2002.

[28] M. Dambacher, W. Eichinger, K. Theisen, and A. Frey, "RT and systolic blood pressure variability after sympathetic stimulation during positive tilt in healthy volunteers," in *Proceedings of Computers in Cardiology*, Sept. 1994, pp. 573–576.

[29] G. Andrassy, A. Szabo, G. Ferencz, Z. Trummer, E. Simon, and A. Tahy, "Mental stress may induce QT-interval prolongation and T-wave notching," *Annals of Noninvasive Electrocardiology*, vol. 12, no. 3, pp. 251–259, 2007.

[30] A. Folino, G. Buja, P. Turrini, O. L., and A. Nava, "The effects of sympathetic stimulation induced by mental stress on signal averages electrocardiogram," *Int. Journal of Cardiology*, vol. 48, pp. 279–285, 1995.

[31] A. Szabo, "The combined effects of orthostatic and mental stress on heart rate, T-wave amplitude, and pulse transit time," *European Journal of Applied Physiology*, vol. 67, no. 6, pp. 540–544, 1993.

[32] H. Scher, J. Furedy, and R. Heslegrave, "Phasic T-wave amplitude and heart rate changes as indices of mental effort and task incentive," *Psychophysiology*, vol. 21, no. 3, pp. 326–333, 1984.

[33] R. Sinha, W. Lovallo, and O. Parsons, "Cardiovascular differentiation of emotions," *Psychosomatic Medicine*, vol. 54, pp. 422–435, 1992.

[34] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: a new approach in human identification," *IEEE Trans. on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.

[35] M. Kyoso and A. Uchiyama, "Development of an ECG identification system," in *Proceedings of the 23rd Annual International Conference of the Eng. in Medicine and Biology Society*, vol. 4, 2001, pp. 3721 – 3723.

[36] T. W. Shen, W. J. Tompkins, and Y. H. Hu, "One-lead ECG for identity verification," in *Proceedings of the 2nd Conf. of the IEEE Eng. in Med. and Bio. Society and the Biomed. Eng. Society*, vol. 1, 2002, pp. 62–63.

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

[38] R. Palaniappan and S. Krishnan, "Identifying individuals using ECG beats," in *Proceedings of International Conference on Signal Processing and Communications*, Dec. 2004, pp. 569 – 572.

[39] K. S. Kim, T. H. Yoon, J. L., D. J. Kim, and H. S. Koo, "A robust human identification by normalized time-domain features of Electrocardiogram," in *Proceedings of 27th Annual Int. Conf on Eng. in Medicine and Biology Society*, Jan. 2005, pp. 1114 –1117.

[40] S. Saechia, J. Koseeyaporn, and P. Wardkein, "Human identification system based ECG signal," in *TENCON, IEEE region 10 Conference*, Nov. 2005, pp. 1 –4.

[41] Z. Zhang and D. Wei, "A new ECG identification method using Bayes' teorem," in *TENCON, IEEE region 10 Conference*, Nov. 2006, pp. 1 –4.

[42] Y. Singh and P. Gupta, "ECG to individual identification," in *Proceedings of IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, Oct. 2008, pp. 1–8.

[43] O. Boumbarov, Y. Velchev, and S. Sokolov, "ECG personal identification in subspaces using radial basis neural networks," in *Proceedings of IEEE Int. Workshop on Intelligent Data Acquisition and Advanced Computing Systems*, Sept. 2009, pp. 446 –451.

[44] C. M. Ting and S. H. Salleh, "ECG based personal identification using extended Kalman filter," in *Proceedings of 10th International Conference on Information Sciences Signal Processing and their Applications*, May 2010, pp. 774 –777.

[45] N. Venkatesh and S. Jayaraman, "Human electrocardiogram for biometrics using DTW and FLDA," in *Proceedings of 20th International Conference on Pattern Recognition (ICPR)*, Aug. 2010, pp. 3838 –3841.

[46] M. Tawfik, H. Selim, and T. Kamal, "Human identification using time normalized QT signal and the QRS complex of the ECG," in *Proceedings of 7th International Symposium on Communication Systems Networks and Digital Signal Processing*, July 2010, pp. 755 –759.

[47] K. Plataniotis, D. Hatzinakos, and J. Lee, "ECG biometric recognition without fiducial detection," in *Proceedings of Biometrics Symposiums (BSYM)*, Baltimore, Maryland, USA, Sept. 2006.

[48] G. Wübbeler, M. Stavridis, D. Kreiseler, R. Bousseljot, and C. Elster, "Verification of humans using the electrocardiogram," *Pattern Recogn. Lett.*, vol. 28, no. 10, pp. 1172–1175, 2007.

[49] G. G. Molina, F. Bruekers, C. Presura, M. Damstra, and M. van der Veen, "Morphological sythesis of ECG signals for person authentication," in *Proceedings of 15th European Signal Proc. Conf.*, Poland, Sept. 2-7 2007.

[50] A. Chan, M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *Instrumentation and Measurement, IEEE Transactions on*, vol. 57, no. 2, pp. 248 –253, Feb. 2008.

[51] C. C. Chiu, C. Chuang, and C. Hsu, "A novel personal identity verification approach using a discrete wavelet transform of the ECG signal," in *Proceedings of International Conference on Multimedia and Ubiquitous Engineering*, April 2008, pp. 201 –206.

[52] S. Fatemian and D. Hatzinakos, "A new ECG feature extractor for biometric recognition," in *Proceedings of 16th International Conference on Digital Signal Processing*, July 2009, pp. 1 –6.

[53] I. Odinaka, P.-H. Lai, A. Kaplan, J. O'Sullivan, E. Sirevaag, S. Kristjansson, A. Sheffield, and J. Rohrbaugh, "Ecg biometrics: A robust short-time frequency analysis," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, Dec. 2010, pp. 1 –6.

[54] C. Ye, M. Coimbra, and B. Kumar, "Investigation of human identification using two-lead electrocardiogram (ECG) signals," in *Proceedings of 4th Int. Conf. on Biometrics: Theory Applications and Systems*, Sept. 2010, pp. 1 –8.

[55] D. Coutinho, A. Fred, and M. Figueiredo, "One-lead ECG-based personal identification using Ziv-Merhav cross parsing," in *Proceedings of 20th Int. Conf. on Pattern Recognition*, Aug. 2010, pp. 3858 –3861.

[56] N. Ghofrani and R. Bostani, "Reliable features for an ECG-based biometric system," in *Proceedings of 17th Iranian Conference of Biomedical Engineering*, Nov. 2010, pp. 1 –5.

[57] M. Li and S. Narayanan, "Robust ECG biometrics by fusing temporal and cepstral information," in *Proceedings of 20th International Conference on Pattern Recognition*, Aug. 2010, pp. 1326 –1329.

[58] P. Ekman, R. Levenson, and W. Friesen, "Autonomic nervous system activity distinguishes among emotions," *Science*, vol. 221, pp. 1208–1210, Sept. 1983.

[59] G. Miller, D. Leven, and M. Kozak, "Individual differences in imagery and the psychophysiology of emotion," *Cognition and Emotion*, vol. 1, pp. 367–390, 1987.

[60] A. Pecchinenda and C. Smith, "The affective significance of skin conductance activity during a difficult problem-solving task," *Cognition and Emotion*, vol. 10, pp. 481–504, Sept. 1996.

[61] R. Picard, "Affective computing: Challenges," *International Journal of Human-Computer Studies*, vol. 59, pp. 55–64, 2003.

[62] E. Vyzas and R. Picard, "Affective pattern classification," in *Proceedings of Emotional and Intelligent: The Tangled Knot of Cognition*, 1998, pp. 176–182.

[63] J. Riseberg, J. Klein, R. Fernandez, and R. Picard, "Frustrating the user on purpose: using biosignals in a pilot study to detect the user's emotional state," in *Proceedings of Computer Human Interaction 98, Conference summary on Human factors in computing systems*. New York, NY, USA: ACM, 1998, pp. 227–228.

[64] F. Nasoz, C. Lisetti, K. Alvarez, and N. Finkelstein, "Emotion recognition from physiological signals for user modeling of affect," in *Proceedings of the 9th Int. Conf. on User Model*, Pittsburg, June 2003.

[65] F. Nasoz, K. Alvarez, C. Lisetti, and N. Finkelstein, "Emotion recognition from physiological signals using wireless sensors for presence technologies," *Int. Journal of Cognition, Technology and Work. Special issue on Presence*, vol. 6, no. 1, pp. 4–14, Feb. 2004.

[66] J. Gross and R. Levenson, "Emotion elicitation using films," *Cognition and Emotion*, vol. 9, pp. 87–108, 1995.

[67] K. Kim, S. Bang, and S. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and Biological Engineering and Computing*, vol. 42, no. 3, pp. 419–427, May 2004.

[68] P. Broersen, "Facts and fiction in spectral analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 4, pp. 766–772, Aug. 2000.

[69] F. Hönig, A. Batliner, and E. Nöth, "Real-time Recognition of the Affective User State with Physiological Signals," in *Proceedings of the 2nd Int. Conf. on Affective Computing and Intelligent Interaction, Proceedings of the Doctoral Consortium*, R. Cowie and F. de Rosis, Eds., 2007, pp. 1–8.

[70] C. Lee, S. Yoo, Y. Park, K. NamHyun, J. KeeSam, and L. ByungChae, "Using neural network to recognize human emotions from heart rate variability and skin resistance," *27th Annual Int. Conf. of the Eng. in Medicine and Biology Society*, pp. 5523–5525, Jan. 2005.

[71] J. Anttonen and V. Surakka, "Emotions and heart rate while sitting on a chair," in *Proceedings of the Conf. on Human factors in computing systems, CHI 05*, 2005, pp. 491–499.

[72] P. Lang, M. Bradley, and B. Cuthbert, "Int. affective picture system (IAPS): Instruction manual and affective ratings." *The center for research in psychophysiology, Technical Report A-5*, 2001, university of Florida, Gainesville, Florida.

[73] A. Nakasone, H. Prendinger, and M. Ishizuka, "Emotion recognition from electromyography and skin conductance," in *Proceedings the 5th Int. Workshop on Biosignal Interpretation*, 2005, pp. 219–222.

[74] M. Benovoy, J. Cooperstock, and J. Deitcher, "Biosignals analysis and its application in a performance setting - towards the development of an emotional-imaging generator," in *Proceedings of the 1st International Conference on Biomedical Electronics and Devices, BIOSIGNALS*, Madeira, Portugal, 2008, pp. 253–258.

[75] R. Mandryk and K. Inkpen, "Physiological indicators for the evaluation of co-located collaborative play," in *Proceedings of the ACM conference on Computer supported cooperative work (CSCW04)*. ACM, 2004, pp. 102–111.

[76] R. Mandryk, S. Atkins, and K. Inkpen, "A continuous and objective evaluation of emotional experience with interactive play environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 06)*. New York, NY, USA: ACM Press, 2006, pp. 1027–1036.

[77] R. Mandryk, "Evaluating affective computing environments using physiological measures," in *Proceedings of Innovative Approaches to Evaluating Affective Interfaces, at CHI 2005*, Portland, USA, April 2005.

[78] R. Mandryk and M. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies," *Int. Journal of Human Computer Studies*, vol. 65, no. 4, pp. 329–347, 2007.

[79] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," in *Proceedings of Affective Dialogue Systems*, 2004, pp. 36–48.

[80] C. Jones and T. Troen, "Biometric valence and arousal recognition," in *Proceedings of the 19th Australasian Conf. on Computer-Human Interaction*, New York, 2007, pp. 191–194.

[81] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, S. M., and M. Schroder, "FEELTRACE: An instrument for recording perceived emotion in real time," in *Proc. of iSCA Workshop Speech and Emotion.*, 2000, pp. 19–24.

[82] "The HeartID system." http://www.comm.utoronto.ca/ biometrics/medical/HeartID.pdf.

[83] F. Agrafioti and D. Hatzinakos, "Fusion of ECG sources for human identification," in *Proceedings of 3rd Int. Symp. on Communications Control and Signal Processing*, Malta, March 2008, pp. 1542–1547.

[84] W. Sethares and T. Staley, "Periodicity transforms," *IEEE Trans. on Sig. Proc.*, vol. 47, no. 11, pp. 2953–2964, Nov. 1999.

[85] D. Mucke, *Elektrokardiographie systematisch.* UniMed Verlag, 1996.

[86] R. Picard, *Affective Computing.* MIT Press, July 2000.

[87] R. Mandryk and T. W. Inkpen, K.M. Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies," *Behaviour and Information Technology*, vol. 25, no. 2, pp. 141–158, March-April 2006.

[88] F. Agrafioti, F. Bui, and D. Hatzinakos, "Medical biometrics: The perils of ignoring time dependency," in *Proceedings of 3rd Int. Conf. on Biometrics: Theory, Applications, and Systems, 2009. BTAS '09*, Washington, Sept. 2009, pp. 1 –6.

[89] C. Zong and M. Chetouani, "Hilbert-Huang transform based physiological signals analysis for emotion recognition," in *Proceedings of Int. Symposium on Signal Processing and Information Technology (ISSPIT)*, Dec. 2009, pp. 334 –339.

[90] N. Huang, Z. Shen, R. Long, M. Wu, Q. Zheng, N. Yen, and C. Tung, "The empirical mode decomposition and hilbert spectrum for nonlinear and nonstationary time series analysis," *Proc. Roy. Soc. London*, vol. 454, p. 903995, 1998.

[91] G. Rilling, P. Flandrin, P. Gonalves, and J. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 936 –939, Dec. 2007.

[92] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. on Biomedical Eng.,*, vol. 32, no. 3, pp. 230 –236, March 1985.

[93] P. McSharry, G. Clifford, L. Tarassenko, and L. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. on Biomedical Eng.,*, vol. 50, no. 3, pp. 289 –294, March 2003.

[94] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Comput. Biol. Med.*, vol. 38, pp. 1–13, Jan. 2008.

[95] A. Arafat and K. Hasan, "Automatic detection of ECG wave boundaries using empirical mode decomposition," in *Proceedings of Int. Conf. on Acoustics Speech and Signal Processing*, April 2009, pp. 461–464.

[96] A. Nimunkar and W. Tompkins, "R-peak detection and signal averaging for simulated stress ECG using EMD," in *Proceedings of 29th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, Aug. 2007, pp. 1261 –1264.

[97] P. Flandrin, P. Goncalves, and G. Rilling, "Detrending and denoising with empirical mode decompositions," in *Proceedings of the European Signal Proc. Conf.*, Sept. 2004.

[98] M. K. L. Molla, T. Tanaka, T. M. Rutkowski, and A. CiChocki, "Separation of EOG artifacts from EEG signals using bivatiate EMD," in *Proceedings of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2010*, March 2010, pp. 562–565.

[99] H. Rau, "Responses of the T-wave amplitude as a function of active and passive tasks and beta-adrenergic blockade," *Psychophysiology*, vol. 28, no. 2, pp. 231–239, 1991.

[100] P. J. Lang, "The emotion probe. studies of motivation and attention." *The American Psychologist*, vol. 50, no. 5, pp. 372–385, May 1995.

[101] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067 –2083, Dec. 2008.

[102] D. Rowe, J. Sibert, and D. Irwin, "Heart rate variability: indicator of user state as an aid to human-computer interaction," in *Proceedings of the SIGCHI Conf. on Human factors in computing systems*, 1998, pp. 480–487.

[103] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. Lim, and J. Suri, "Heart rate variability: a review," *Medical and Biological Engineering and Computing*, vol. 44, pp. 1031–1051, 2006.