The Use of EEG Signals

For Biometric Person Recognition

A Thesis Submitted to the University of Kent

For the Degree of Doctor of Science

In Electrical/Electronic Engineering

By

Su Yang

September 2015

Abstract

This work is devoted to investigating EEG-based biometric recognition systems. One potential advantage of using EEG signals for person recognition is the difficulty in generating artificial signals with biometric characteristics, thus making the spoofing of EEG-based biometric systems a challenging task. However, more works needs to be done to overcome certain drawbacks that currently prevent the adoption of EEG biometrics in real-life scenarios: 1) usually large number of employed sensors, 2) still relatively low recognition rates (compared with some other biometric modalities), 3) the template ageing effect.

The existing shortcomings of EEG biometrics and their possible solutions are addressed from three main perspectives in the thesis: pre-processing, feature extraction and pattern classification. In pre-processing, task (stimuli) sensitivity and noise removal are investigated and discussed in separated chapters. For feature extraction, four novel features are proposed; for pattern classification, a new quality filtering method, and a novel instance-based learning algorithm are described in respective chapters. A self-collected database (Mobile Sensor Database) is employed to investigate some important biometric specified effects (e.g. the template ageing effect; using low-cost sensor for recognition).

In the research for pre-processing, a training data accumulation scheme is developed, which improves the recognition performance by combining the data of different mental tasks for training; a new wavelet-based de-noising method is developed, its effectiveness in person identification is found to be considerable. Two novel features based on Empirical Mode Decomposition and Hilbert Transform are developed, which provided the best biometric performance amongst all the newly proposed features and other state-of-the-art features reported in the thesis; the other two newly developed wavelet-based features, while having slightly lower recognition accuracies, were computationally more efficient. The quality filtering algorithm is designed to employ the most informative EEG signal segments: experimental results indicate using a small subset of the available data for feature training could receive reasonable improvement in identification rate. The proposed instance-based template

reconstruction learning algorithm has shown significant effectiveness when tested using both the publicly available and self-collected databases.

Acknowledgements

I would like to express my sincere thanks to my supervisor Dr Farzin Deravi for the continuous support of my PhD study and research, for his patience, immense knowledge and enthusiasm. His guidance helped me a lot during the research and writing of this thesis.

Besides my supervisor, I would like to thank the rest of my PhD panel members: Professor Mike C. Fairhurst and Dr Sanaul Hoque, for their encouragement, insightful comments and hard questions.

I thank my fellow lab-mates in the Intelligent Interactions Group: particularly Shivam and Richard, for the inspiring discussions and for all the fun we have had in the last four years. Also I thank my manager Saritha Arunkumar during my internship in IBM. In particular, I am grateful to Dr Graham Bent for his enormous patience and support for my work in IBM.

Last but not the least: I would like to thank my parents, without their support I could never reach this point.

Declaration

I certify that I have read and understood the entry in the Project Handbook on Plagiarism and Duplication of Material, and that all material in this Dissertation is my own work, except where I have indicated with appropriate references.

Signed.....

Date.....

TABLE OF CONTENT

ABBREVIATION LIST

ANN	Artificial Neural Network
AR	Models Autoregressive Models
BCI	Brain Computer Interaction
CMC	Cumulative Match Characteristic
CRR	Correct Recognition Rate
DBI	Davies Bouldin Index
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
ECG	Electrocardiography
ED	Euclidean Distance
EEG	Electroencephalographic/Electroencephalography
EER	Equal Error Rate
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
ERP	Event Related Potential
FAR	False Acceptance Rate
FRR	False Rejection Rate
GMM	Gaussian Mixture Model
HHT	Hilbert-Huang Transform
HT	Hilbert Transform
HTER	Half Total Error Rate
ICA	Independent Component Analysis
IMF	Intrinsic Mode Functions

InsAmp	Instantaneous Amplitude
InsFreq	Instantaneous Frequency
ITR	Instance-based Template Reconstruction
***k*-NN**	*k*-Nearest Neighbour
LDA	Linear Discriminate Analysis
LDC	Linear Discriminate Classifier
MFCC	Mel-frequency Cepstral Coefficients
PCA	Principal Component Analysis
PSD	Power Spectral Density
RBF	Radial Based Function
SampEn	Sample entropy
SD	Standard Deviation
SNR	Signal-to-Noise Ratios
SVM	Support Vector Machine
TAR	True Acceptance Rate
VEP	Visual Evoke Potential
WPD	Wavelet Packet Decomposition
WT	Wavelet Transform

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is devoted to an investigation of the use of electroencephalographic (EEG) signals for biometric person recognition. The structure of this chapter is formed as follows: Section 1.1 provides a background to biometrics person recognition and the motivation for the use of EEG signals for biometric applications. The structure of the thesis is described in Section 1.2. The limitations of this thesis are listed in Section 1.3. Section 1.4 lists the publications involved in this thesis, including the published works and the papers under review.

1.1 Background & Motivation

The traditional ways of access control include using token-based identification systems, such as a driver's license or passport, and knowledge-based identification systems, such as a password or personal identification number [1]. Biometric person recognition, however, relies on human physiological and/or behavioural characteristics for establishing identity. A range of such distinctive, measurable characteristics can be used have been successfully used to recognize individuals. Biometric modalities/identifiers are considered to be unique to individuals: they may be more reliable in verifying identity than token and knowledge-based methods [1].

Biometric characteristics are often categorized as being of physiological or behavioural nature [2]. Physiological characteristics usually are related to the shape/colour of the body. Examples include, but are not limited to fingerprint, palm veins, face recognition, DNA, palm print, hand geometry, iris and retina recognition. Behavioural characteristics are related to the behavioural patterns of a person, including but not limited to typing rhythm, gait, and voice [2].

With the development of the low-cost EEG sensor devices, using EEG signals for biometric recognition has been drawing attentions from researchers. Because of the

continuous monitoring characteristic, EEG signal as a biometric modality may be difficult to be spoofed.

The first human EEG signal was recorded by a German physiologist and psychiatrist Hans Berger (1873–1941) in 1924 [3]. Since then, EEG has been conventionally employed for clinical usages. However, EEG signals may be a controversial biometric modality due to their relatively late appearance in the biometrics field. EEG signals elicit from the brain, which is a physiological organ with sophisticated inner structures, which makes it a physiological-based modality. On the other hand, the EEG signal is considered non-stationary [4]; therefore, the impact of template ageing effect of EEG as a biometric modality may be considerable. In this thesis, the EEG signals are considered as one type of behavioural biometric modality.

The use of EEG signals for biometrics was first proposed in 1999 by Poulos *et al*. [5] [6]. There have been around one hundred papers published in this area, and much progress has been made since then. However, as a relatively new biometric modality, there are still a lot of issues yet to be investigated before its deployment in real-life scenarios. Some of the important issues include: developing new features with better recognition performance using traditional (medical) sensor(s) for data capture; improving the performance while using low-cost (less intrusive) sensor systems for data collection; investigating and alleviating the template ageing effects, which may be quite considerable in EEG-based biometric recognition. Given these existing gaps in using EEG for biometric recognition, additional efforts need to be enforced and a systematic investigation of the EEG biometrics becomes a necessity.

1.2 Thesis Structure

There are ten chapters in this thesis: eight research topics are covered in eight chapters excluding the Introduction and the Conclusion chapters. The research topics are described as follows:

1) In the literature review chapter (Chapter 2), usability-related factors are investigated based on an extensive survey of existing literature in using EEG signals for biometric applications.

2) Due to the inadequacies of existing EEG databases for the evaluation of biometric systems, a new database is collected. The data collection protocol for this database is presented in Chapter 3.
3) The results of a series of preliminary investigations are reported in Chapter 4, aimed at establishing the initial parameters of the EEG-based biometric systems that will be used in the rest of this work.
4) The impact of different types of sensory activities on EEG signals for biometric recognition is experimentally investigated in Chapter 5 with the purpose of finding the optimal stimulus in triggering the production of informative signals. This investigation relies on the use of a publicly available database with six types of task-related EEG data.
5) Chapter 6 covers the feature extraction, where four new features are developed and their effectiveness experimentally evaluated.
6) The quality of the resulting features after feature extraction has been found to vary substantially. Chapter 7 a quality filtering algorithm is developed aimed at preserving the EEG segments with relatively high quality before the feature extraction.
7) In Chapter 8, a novel instance-based template reconstruction learning algorithm is developed for classification. One major motivation for developing the new algorithm is to alleviate the template ageing effects of EEG biometrics, while using low-cost sensor for data collection.

The overall thesis structure is illustrated by Fig.1.1. Each chapter is devoted to addressing its corresponding research questions; the bold blocks indicate the chapters which contain original contributions.

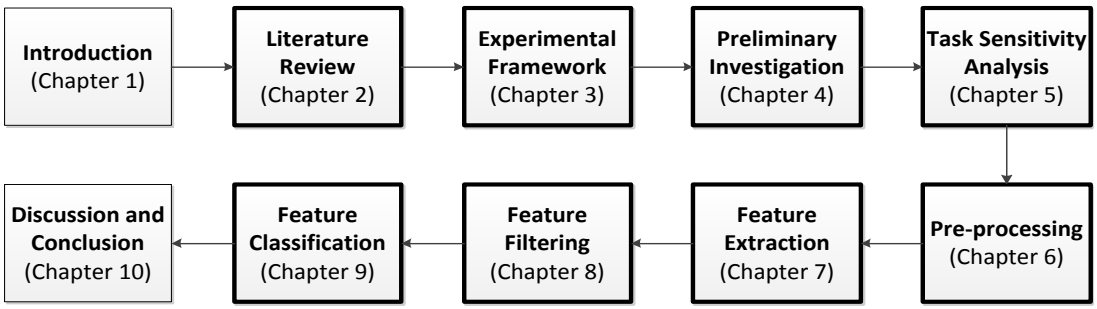


Fig. 1.1 Thesis structure

1.3 Contributions

There are three major contributions in the thesis: 1) a new database was created particular for EEG biometrics recognition; 2) several new features were developed to improve the biometric system; 3) novel classification algorithms which led considerable boost in recognition performance.

The new database is designed to address some critical issues in the EEG biometric field. The publicly available databases used for biometrics do not have the potential for researchers to investigate the template consistency/changing issue, however, the EEG signals as biometric templates are found inconsistent over time. It is, therefore, a big obstacle in the path of its practical implementation.

Four new features are presented in the thesis. Two wavelet-based features and two features based on improved empirical mode decomposition (EMD). One of the EMD-based features provided the best biometrics recognition performance [7]; the wavelet-based features, however, indicated much faster processing speed than the EMD-based features during the feature extraction stage. It needs to note that the inefficiency of the EMD-based method mainly stem from its recurrent sifting process, the optimization of the programming is only a minor negative factor.

A novel instance-based template reconstruction algorithm is developed and tested using multiple EEG databases. The proposed algorithm is particularly advanced in dealing with the temple consistency/changing effects over time. The experimental results obtained from the self-collected database showed significant improvements.

1.4 Limitations

One limitation of the specially-designed database is the number of subjects included for data capture using the low-cost sensor system. The self-collected database contains 27 subjects with recordings from multiple sessions made on different days, which may still not be large enough to justify the statistical significance of the obtained results. Furthermore, due to the quality of the data from the low-cost sensor, it can be predicted that with the number of subjects further increasing, the recognition

performance of the EEG biometric system tends to reduce. The next chapter is devoted to the literature review of the relevant research.

1.5 Publications

The following papers have been published based on the work reported in the thesis:

- Su Yang, and Farzin Deravi, "On the Effectiveness of EEG Signals as a Source of Biometric Information", *Proc. Third Int'l Conf. Emerging Security Technologies (EST)*, pp.49-52, 2012.
- Su Yang, and Farzin Deravi, "Wavelet-Based EEG Pre-processing for Biometric Applications", *Proc. Fourth Int'l Conf. Emerging Security Technologies (EST)*, pp. 43-46, 2013.
- Su Yang, and Farzin Deravi, "Quality Filtering of EEG Signals for Enhanced Biometric Recognition", *Int'l Conf. Biometrics Special Interest Group (BIOSIG)*, vol. 212, pp. 201-208, 2013.
- Yang, Su, and Farzin Deravi, "Novel HHT-Based Features for Biometric Identification Using EEG Signals." In *22nd International Conference on Pattern Recognition (ICPR)*, pp. 1922-1927, IEEE, 2014.
- Primary inventor of an US patent on an advanced instance-based generic learning algorithm for pattern recognition, using EEG signal as the embodiment, collaborated with IBM UK. Docket Number *GB920150042US1*, Patent Title: *Feature Selection Algorithm Under Conditions of Noisy Data and Limited Recording*.

Chapter 2

Previous Work on EEG Signal Acquisition and Analysis

Electroencephalography (EEG) is a recording of the electrical activities of the brain along the scalp, which is generally considered as a non-invasive data collection method. EEG measures the voltage fluctuations resulting from ionic current within the neurons of the brain [8], therefore, it may be employed to measure some inner characteristics of the brain. One of the earliest investigations on the brain-related functions is monitoring the electroencephalographic signals of animals (rabbits' and monkeys' brains, in 1890) [9][10]. The first human EEG was recorded in 1924 [3], later the EEG gradually became a useful tool in diagnosing brain disease (epileptic seizures [11], for instance). Considering the physiological differences between individual's brains, EEG signals may be expected to possess the potential to not only indicate the brain's functions, but also dissimilarities between individuals as manifested by the electrical activity of their brains. Such an assumption has led to attempts at developing works on EEG-based person recognition systems, where EEG signals are used as a biometric modality.

The EEG signal has been employed as a biometric modality by a number of researchers in the laboratory environment (for example in [6]); however, many problems still remain and need to be addressed before considering its application in real-life scenarios. The main goal of this investigation is to extensively review the existing research in using EEG signals for biometric recognition. It also highlights the current challenges which potentially hinder the wider deployment of EEG biometrics, mainly surrounding the issue of practicality and usability.

Some useful survey papers reviewed the existing EEG biometric works focused on different angles. Study in [12] addresses the EEG biometric reports from the theoretical aspects, including brain rhythms and elicitation protocols. Another paper [13] focuses on reviewing the state-of-the-art methods and future perspectives of using EEG for biometric applications. This work focus on investigating the usability

aspects of the EEG-based biometric systems, also propose some suggestions to improve the system performance in the real-life environment.

This chapter is organized into four major sections: a review of the sensors and stimulus used to trigger and acquire EEG signals for person recognition is presented in Section 2.1. In Section 2.2 a survey of the techniques used in generating features for EEG biometrics is provided. Section 2.3 describes the relevant techniques in EEG biometrics for classification. Section 2.4 first explores a series of factors that affect the usability of an EEG biometric system, followed by introducing a metric which combines these factors to provide a better measurement of the usability of the systems.

Section 2.2 and Section 2.3 represent the main stages in EEG biometrics analysis: Section 2.2 includes both signal pre-processing (such as temporal segmentation) and feature generation. Section 2.3 covers feature selection and classification stages. In Section 2.4, a series of comparisons of the published systems using the new metric are presented, which are grouped by different feature extraction techniques. Template ageing effects in EEG biometrics is also included in this section, as it affects the ultimate effectiveness and usability of EEG biometric systems. Conclusions are presented in Section 2.5.

The detailed structure of this chapter is illustrated by Fig. 2.1. The signal acquisition section is divided into three subsections: resting, sensory and cognitive. Feature extraction/generation section is further divided into five subsections based on the techniques reported in the literature: Subsection 2.2.1 is devoted to features based on the Fourier Transform; feature extraction based on Autoregressive (AR) Models is included in Subsection 2.2.2; Subsection 2.2.3 is related to Wavelet Transform methods; the Hilbert-Huang Transform (HHT) and its related methods for feature extraction are described in Subsection 2.2.4; in Subsection 2.2.5 other methods are compared. The feature classification section is divided into four subsections: *k*-Nearest Neighbour methods are covered in Subsection 2.3.1; Subsection 2.3.2 is devoted to the classification using some linear classifiers. EEG signal classification using Neural Network methods are included in Subsection 2.3.3; Kernel methods and their application in EEG biometrics are described in Subsection 2.3.4. A new metric

which contains four usability-oriented factors along with the recognition accuracy is then proposed to evaluate the overall usability of an EEG biometric system.

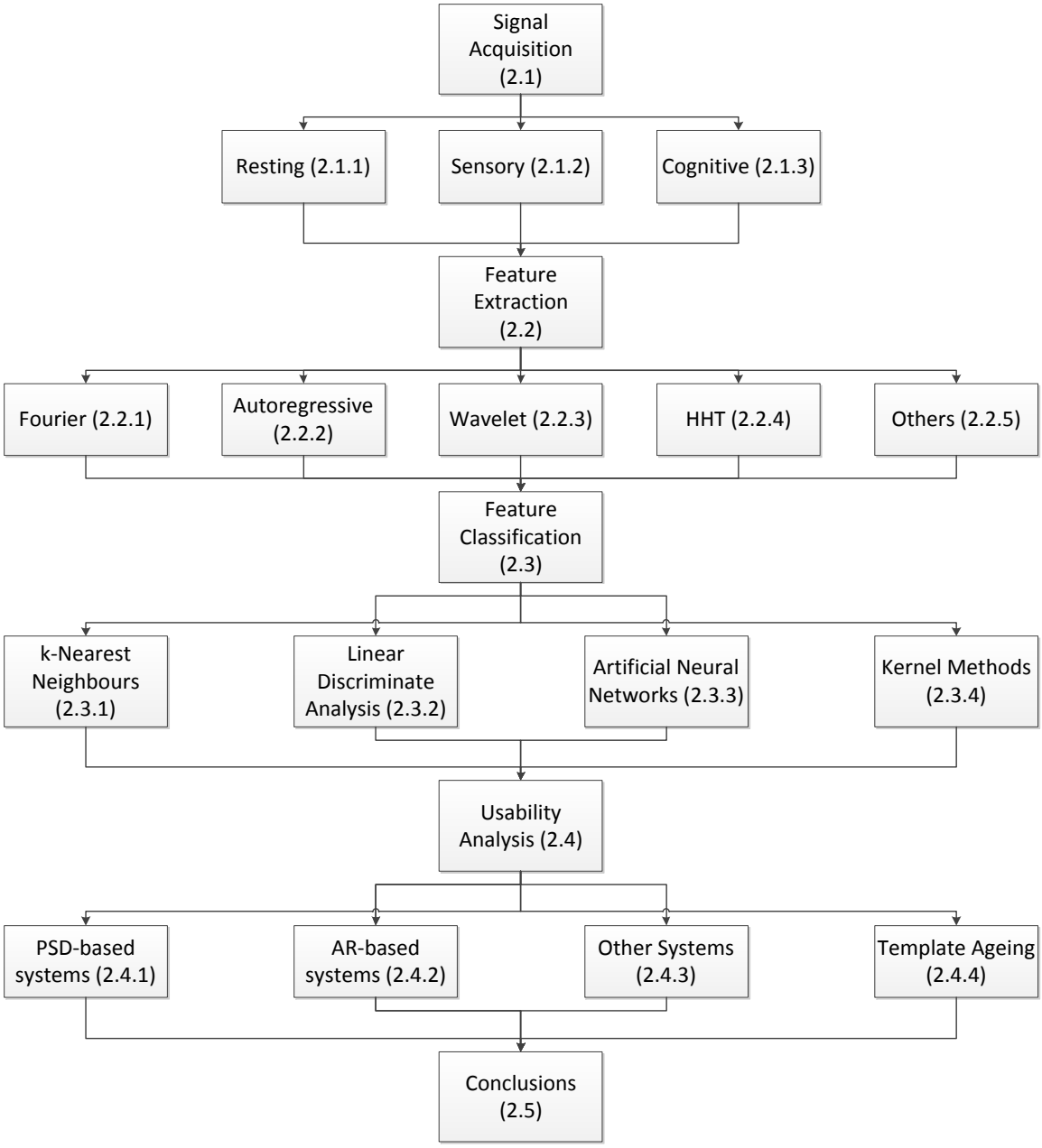


Fig. 2.1 Structure of this chapter

2.1 Signal Acquisition

Two types of sensor system have been used for biometric purposes: a) medical-grade sensor systems and b) low-cost sensor systems. The medical-grade systems

conventionally contain large number of electrodes for data capture; the sensors usually need to be moistened by electrolytes, such as saline solution. Using such sensor system can obtain signals with much better quality, but the deployment of the wet sensors could become a shortcoming in EEG biometric applications. The distribution of the electrodes over the scalp conventionally follow the 10-20 system [14] or 10-10 system [15]. The 10-20 system and 10-10 system are internationally recognized methods to describe and apply the location of electrodes over the scalp in the context of an EEG test or experiment. The "10" and "20" refer to the fact that the actual distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull [16]. The 10-10 system follows the same notion as 10-20 system, but with higher density of the electrode distribution over the scalp. On the other hand, low-cost sensor systems contain a very small amount of electrodes (even single electrode), using the dry sensor(s) to reduce the cost and improve the usability of the system. Fig. 2.2 highlights the data capture characteristics of these two data capture systems.

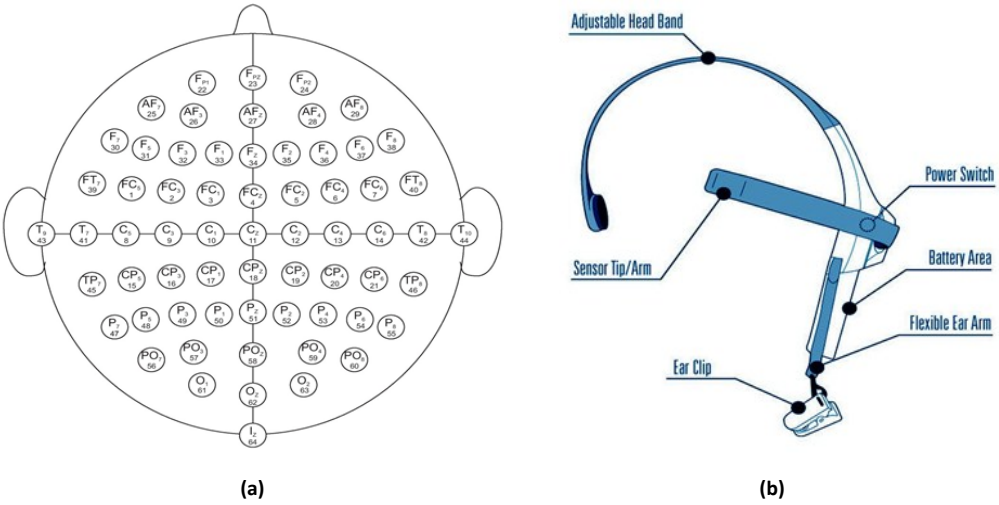


Fig. 2.2 (a) A 64 electrodes positioning map based on 10-10 system [15]; (b) A single Fp1 electrode low-cost headset [17].

As a relatively new biometric modality, many sensor systems used in the research for EEG person recognition are directly those used in the medical field. Therefore, the number of electrodes usually is large and the sensors are scattered all over the scalp [14]. Sensor systems that have been used in biometrics research include, for example,

the BCI2000 system [18][19]. On the other hand, to be considered as a successful biometric modality, good recognition performance is a necessity but may not be enough: a sensor system with less demanding characteristics are equally important in the application. Researchers have, therefore, been testing the performance of low-cost sensor headsets for biometric recognition. Popular headsets such as Emotiv [20] and Neurosky products [17] have been tested in biometric scenarios. Emotiv EPOC is the first low-cost sensor system that using electrolyte [20]; the chips of Neurosky products have been used in many inexpensive EEG collection devices. Table 2.1 shows some commercial low-cost EEG sensor systems, which are or potentially could be used for biometric person recognition.

Table 2.1 List of market available low-cost EEG headsets

Device	**Electrode(s)**	**Sensor type**	**Released**
NeuroSky MindSet [21]	1	dry	2007
Neural Impulse Actuator [22]	3	dry	2008
Emotiv EPOC [20]	14	wet	2009
Mindflex (Uses NeuroSky chips)	1	dry	2009
MindWave [21]	1	dry	2011
XWave headset (NeuroSky chips)	1	dry	2011
Melon Headband [23]	4	dry	2014
HiBrain [24]	1	dry	2014
iFocusBand [25]	1	dry	2014
Muse [26]	4	dry	2014
OpenBCI [27]	8 or 16	dry/wet	2014
Aurora Dream Headband [28]	1	dry	2015
Emotiv Insight [20]	5	dry	2015

The choice of EEG sensor(s) and their placement have a significant impact on the performance and usability of EEG biometric recognition systems. It worth mentioning the sensitive frequency ranges of EEG signal. The typical EEG bands are traditionally divided into several bands, namely: delta (< 4 Hz), theta (4 Hz – 7 Hz), alpha (8 Hz – 15 Hz), beta (16 Hz – 31 Hz) and gamma (> 32 Hz) [29]. Usually, only the frequency below 50 Hz is taken into consideration in EEG biometric recognition.

An essential aspect to EEG data acquisition for biometric person recognition is the state of the brain and the nature of its activity. The identity information bearing

signals that are produced by the brain may crucially depend on the type of brain activity in which the user is engaged. Often there may need to be a predefined stimulus to evoke the desired brain activity for producing consistent results. The choice of stimulus that would result in optimal production of identity-bearing signals for biometric person recognition purposes is still a subject of research. In the following sections, some relevant published approaches are surveyed covering three types of brain stimulation: resting state, sensory (audio/visual) stimuli and cognitive tasks - (verbal instructions).

2.1.1 Resting State

For the data acquisition process during the resting state, usually the subject is asked to sit on a chair in a quiet environment, either with eyes open or closed. The resting state scheme is the least demanding in terms of the need for additional equipment to generate an external stimulus: the users passively providing EEG signals without extra instruction during the data collection.

Su *et al*. [30] used only the Fp1 position (frontal region) electrode for data collection from 40 healthy subjects while resting on a sofa with their eyes closed. Each subject provided 60 minutes of recordings (12 separated sessions) in total and half of this data was randomly selected for training and the rest was used for testing. A correct recognition rate (CRR) of 97.5% was reported.

Rocca *et al*. [31] reported person identification performance using two resting state subsets of a relatively large (108 subjects) publicly available EEG database: one with eyes open and the other with eyes closed. Of the one minute's EEG recording for each subset per subject, data of ten seconds were used for testing and the rest for training. Performance of 100% CRR was achieved by the fusion of conventional power spectral features and a functional connectivity feature that they proposed. Fraschini *et al*. [32] employed the same dataset for a verification scenario, an Equal Error Rate (EER) of 4.4% was obtained using the "eigenvector centrality" features extracted from gamma band (30 Hz-50 Hz).

A potential obstacle of using the resting state EEG data for biometrics recognition may be the ambiguity of the instruction for users, which may be interpreted by the subject in different ways and resulting in incommensurable data.

2.1.2 Sensory Stimuli

An event-related potential (ERP) is one kind of brain response that directly triggered by "a specific sensory, cognitive, or motor event" [33]. The P300 wave is one such ERP component obtained during the process of decision making, such as the reaction to the oddball paradigm [34]. "Research indicated that when P300 signal was recorded by EEG, it surfaces as a positive deflection in voltage with latency (delay between stimulus and response) of roughly 250 to 500 ms" [35]. P300 is one type of visual evoke potential (VEP), which belongs to the ERP category. In order to obtain visual evoked data, subjects often are instructed to watch a series of fast flashing/moving pictures: the EEG data (a matter of milliseconds) after each picture flash is recorded for analysis. One advantage of using the VEP data for biometric recognition is that the expected waveforms could be accurately measured during the feature extraction. For instance, the P300 waveform can be detected at 250 milliseconds to 500 milliseconds after the visual stimulus. Researchers have been using this particular waveform for person recognition.

Palaniappan *et al*. [36] employed the P300 Visual Evoke Potential (VEP) for feature extraction while people were viewing a set of standardized pictures originally proposed in [37]. The experiment comprised of 10 subjects using an EEG cap with 61 electrodes. The maximum identification rate achieved was 95% using the data recorded in a single session (part of the session used for training and the rest for testing). Similar but improved approaches have been tested using databases with larger population leading to the maximum performance of 98.12% with a database of 102 subjects [38][39][40][41][42].

In [41], Davies Bouldin Index (DBI) has been used to filter out the less informative electrodes: 35 out of 61 electrodes were selected, tested using a database of 40 subjects. The gamma band VEP data was extracted using an Elman neural network

for the system training: a maximum recognition rate of 98.56±1.87% was reported. It was mentioned in this work that, though the performance would have been further improved by utilizing all the electrodes' data, the increased complexity of deploying more electrodes is not justified for about 1% improvement in accuracy that will result.

One possible drawback of using visual stimuli for biometric applications is the need for external devices to trigger the VEP signals. This may make the resulting biometric system more complex compared with alternatives based on using the resting state or directed cognitive activity.

2.1.3 Cognitive Activities

Considering the limitations of the resting state and sensory stimulus approaches for EEG biometric applications, researchers have explored other approaches which may result in brain signals that are more repeatable, and potentially more individually distinctive. The cognitive stimulus involves asking subject(s) to perform various cognitive tasks during the data collection: could be actual movement tasks or imagery tasks. The use of cognitive stimuli was first reported in 2005 when EEG data recorded during the performance of a variety of mental tasks (including: mathematical calculation, geometric figure rotation, mental letter composition and visual counting) were first used in an identification scenario to distinguish 4 subjects [43].

Marcel *et al*. [44] used the data captured while imagining hand movements for a biometric authentication scenario. Power Spectral Density (PSD) features of EEG signals from different subjects were compared using Gaussian Mixture Models (GMM) and, a half total error rate of 7.1% was reported for 9 subjects using 8 electrodes (using data recording of 16 minutes for training and 4 minutes for testing).

Gui *et al*. [45] recorded EEG signals from 30 subjects performing a series of reading tasks. Two types of features were extracted by measuring the Euclidean Distance and Dynamic Time Warping between the EEG signal series of each individual. 55 seconds of recording per subject were obtained. About 10 seconds' data were randomly picked for training and another 20 seconds were used for the test. An

identification rate of 81%, using an electrode in the Oz position, with the Euclidean Distance derived features was employed.

One major issue in employing EEG as a biometric modality has been the complexity of setting up the data acquisition system, especially when high-end medical devices are used for data acquisition. Therefore, the use of low-cost sensors has become an interesting subject for research despite their expected low signal quality. Chuang *et al*. [46] reported a system which employed only a single Fp1 electrode (NeuroSky MindSet [17]); two 40-50 minute data acquisition sessions on separate days were conducted. Different activities were performed, including: Breathing, Eye and Audio Tone, Object Counting *et cetera* (7 tasks). The half total error rate (HTER $= \frac{FAR+FRR}{2}$) of 1.1% for a database of 15 subjects was 99%. However, an identification rate of only 22% was achieved using the same database.

Similar to the EEG signals captured during the resting state, the mental/cognitive task-stimuli also suffer from the ambiguity problem. However, usually the challenge of employing such tasks does not stemmed from the task instruction or user's interpretation, but from the data analysis. It is often difficult (if not impossible) to extract the task-related content from the overall data for analysis and leave behind the part of the brain activity that is not related to the task and does not carry identity information: commonly all of the data acquired during a recording session is analysed regardless the nature of the task.

2.2 Feature Extraction

The choice of features is crucial to the performance of an EEG biometric system. The nature of EEG signals affects the developing of the biometric features: some features reported in the literatures are based on Fourier Transform, which are designed to capture the energy/spectra of the signals; as time series, some features are designed to capture the time-dependent information of the EEG signals; some wavelet-based features are developed to capture both the time and the frequency characteristics of EEG signals.

An extensive review of the published research using different feature extraction techniques for EEG biometrics is provided in this section. Each subsection begins with a brief description of the technique's rationale, followed by the related literature reports.

2.2.1 Power Spectral Density (Fourier Methods)

One of the frequently employed features in EEG biometrics is the Power Spectral Density (PSD). Conventionally, the PSD feature is computed from the Fourier Transform (FT) of the time-domain signal, which indicates the spectral density distribution of the signal in the frequency domain (though it is possible to compute the PSD from the time domain series directly). However, FT is only well-defined when the signal is stationary [47], and many bio-signals such as EEG signals are non-stationary series. Therefore, usually the truncated Fourier Transform $\hat{x}_T(\omega)$ over a finite interval [0, *T*] is computed instead. Within that interval the signal is assumed to be stationary and the PSD $S_x(\omega)$ of the signal $x(t)$may be computed as follows [47], where $\hat{x}_T(\omega)$: is the Fourier Transform of $x(t)$ [47]:

$$S_x(\omega) = \lim_{T\to\infty} E[|\hat{x}_T(\omega)|^2] \tag{2.1}$$

As one of the most popular features in EEG biometrics, PSD was first employed in an identification scenario in 1999. Poulos *et al*. [6] proposed to utilize the spectral information of EEG signals to distinguish persons' identities. Four subjects, each with 45 three-minute recordings were involved in the experiments. Another group of 75 subjects treated as a single class, each with only one recording per subject, was also used for classification against the other four subjects. The voltage between O2 and Oz positions was measured and used in their investigation. The power spectral values of three overlapping frequency bands were used as features: 7-10Hz, 8-11Hz and 9-12Hz. The features were fed into a Learning Vector Quantization (LVQ) network for classification. A series of binary classifications were conducted: data from one class of the four subjects was compared with the 75 subjects-group (as the other class) in the dataset. The correct recognition rate obtained ranging from 80% to

100% across the four subjects. Their experimental results clearly indicated there were certain individual differences amongst subjects.

Palaniappan *et al*. also proposed to use power spectra as the feature for EEG-based identification [48]. They employed the power spectra which centred around 40 Hz (Gamma band) while 10 users were subjected to a series of flashing picture sets containing various objects. The users were asked to view a picture of a single object for each data recording; each subject provided 40 VEP signals (recordings), in a 5.1s time interval. A medical-grade system with 61 electrodes was used for data acquisition. The features were passed on to a fuzzy ARTMAP (FA) classifier and an identification rate of 95% was reported. An interesting aspect of this work was the relatively short training time of about 3 minutes.

PSD features have been used for identification within a larger population of 102 users in [42]. A series of different power spectral features with increased bandwidth (25 to 56 Hz, still within gamma band) were developed using the Multiple Signal Classification (MUSIC) algorithm. Visual stimuli were employed to trigger informative events for biometric recognition. A Manhattan distance-based k-NN classifier was used, reporting a maximum CRR of 98.12%. However, the sensor system contained all the 61 electrodes in order to reach this performance level.

Using EEG-based power spectral features in a verification scenario was explored by Marcel *et al*. in [49]. The power spectral density within the 8-30Hz band from eight centro-parietal channels (C3, Cz, C4, Cp1, Cp2, P3, Pz, and P4) was estimated. The data was collected from 9 subjects, in 12 sessions over 3 days (4 sessions per day). Each session per day lasted 4 minutes with 5 to 10 minutes of break; three tasks were assigned to each subject (imagination of repetitive self-paced left hand movements; imagination of repetitive self-paced right hand movements; generation of words beginning with the same random letter), thus three records per session. A Gaussian Mixture Model (GMM) was employed for modelling the features. Maximum A Posteriori (MAP) classification was used to predict the user identity. A minimum Half Total Error Rate (HTER) of 7.1% was reported.

Safont *et al*. investigated the performance of EEG signals in an authentication scenario for a larger database [50]. Data from two forehead electrodes (Fp1 and Fp2) were used; subjects were sitting in a dark room in a resting state during the data capture. A series of different features including PSD were used for feature extraction. A population of 70 subjects was used for the verification evaluation, 20 of them acted as intruders and the rest were considered genuine. Genuine subjects had data recording durations ranging from 3 to 5 minutes and each intruder provided 2 minutes' data. A relatively good performance of 2.4% EER was reported. However, the data contained only one session per subject and the effect of template ageing could not have been measured.

PSD-based features have also been used for a single sensor system. According to the work reported by Miyamoto *et al*.[51], the data obtained from the Fp1 position was used for analysis while subjects were in a resting state with eyes closed. Variance of the power spectrum (alpha band) was computed and used in a verification scenario. The collected dataset contains 23 subjects; each subject was recorded for 3 minutes in the same day (single session). The reported equal error rate was 21%. The higher error rates compared to some other multi-electrode acquisition systems may be attributed to the substantially reduced biometric information available when only a single electrode is used.

To alleviate this degradation in verification performance, the concavity and convexity of the alpha band spectral distribution were explored in [52]. An improved EER of 11% was achieved using the same database (23 subjects resting status). The novelty of this work lies in not only measuring the short-term energy value of the signal (PSD), but also the overall spectral distribution (shape) of the signal over a period of time as well.

The identification performance of a single EEG electrode was investigated using a database comprising of 40 subjects in [30]. Data from the Fp1 electrode was collected while subjects were in a resting state with eyes closed. The PSD coefficients within selected frequency ranges were used as features (5 Hz to 32Hz). Five separated recordings of about 12 minutes in total for each user were obtained and 50% of all the

data per individual was used for training a *k*-NN classifier. An identification accuracy of 97.5% was reported.

Using low-cost sensors systems for EEG biometrics in conjunction with PSD features has also been explored [46]. The NeuroSky MindSet [17] single sensor system has been used for data collection: different mental tasks were performed by each subject and repeated five times per session. Two sessions of data on separate days (40-50 minutes with multiple tasks per session) were captured; PSD coefficients of alpha band (8-12 Hz) and the beta band (12-30 Hz) were extracted; the cosine similarity was used for classification in the verification scenario. The reported HTER was 1.1% by applying a "customized threshold" [46] for 15 subjects.

The EEG Motor Movement/Imagery Dataset (containing data of four mental tasks and two resting state tasks) has been used for evaluating EEG-based biometric systems [31]. Rocca *et al*. used only its baseline data (resting state of one minute, single session) for identification. A CRR of 100% was reported for a population of 109 subjects by fusing the conventional PSD and a spectral coherence connectivity feature they proposed. However, this data were obtained from a single recording, therefore, using part of this single recording for training and the rest for testing may lead to positively-biased results as it ignores any template ageing/changing effects that may be present, due to the unstable nature of the EEG signals

2.2.2 Autoregressive Model (AR)

AR models have been another very popular feature since the early development of EEG biometrics [5]. An AR model is a time domain representation of a random process. The classic AR model for a random process X_t is defined as follows [53]:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{2.2}$$

where $\varphi_{1,\ldots,}\varphi_p$ are the coefficients of the model, c is a constant and ε_t is the white noise. Expression (2.2) can be equivalently re-written using the following form by employing the lag operator *B*:

$$X_t = c + \sum_{i=1}^{p} \varphi_i B^i X_t + \varepsilon_t \tag{2.3}$$

as such, the signal X_t can be represented by a series of AR coefficients φ_i and white noise ε_t. The AR coefficients may reveal certain intrinsic characteristics of EEG signals, and therefore, suitable candidates as biometric features.

One of the early work using AR coefficients for EEG biometrics is reported by Paranjape *et al*. [54]. In their experiment, data from 40 subjects was obtained from one electrode (at the P4 position). The data recorded for each subject contained 8 epochs (8.533 seconds per epoch) while subjects were in a resting state. A correct classification rate of 80% was achieved using a Linear Discriminate Analysis (LDA) classifier. As an early EEG biometrics paper, the highlight of the experiment was the use of only one electrode for identification; however, the manual contamination removal that was used for data preparation maybe considered as a restriction on its applicability.

Using AR coefficients as features in a verification scenario was also investigated by Riera *et al*. [55]. They reported a True Acceptance Rate (TAR) of 96.6% in a database containing 87 subjects; 36 of which acted as impostors. Data capture was conducted while subjects were in a resting state. The recording length for each subject was between 2 to 4 minutes; the genuine subjects were recorded at 4 sessions in different days whereas the impostors were recoded only once. Only two frontal electrodes (Fp1 and Fp2) were used and the frequencies above 50 Hz were removed. Fisher's discriminant analysis was used for classification and an EER of 3.4% was reported. This result was obtained by using a number of different features (including PSD) and the outputs from separated classifiers were combined by fusing the scores provided by 28 Fisher's discriminant analysis classifiers with different features.

AR coefficients also seem to work fine for other kinds of EEG stimuli. Brigham and Kumar [56] demonstrated that using the EEG signals while subjects were imagining language syllables could achieve successful person identification. The subject imagined two syllables without performing any overt physical actions during the data recording process. The employed data acquisition system contained 128 electrodes

and the data produced by 106 of them were kept for feature extraction. Only six volunteers were involved in the experiment and a CRR of 99.76% was achieved using a Support Vector Machine (SVM) for classification.

Campisi *et al*. investigated the recognition performance Under an "Eyes Closed Resting Conditions" protocol [57]. The data was captured from 48 subjects. The sixth order AR coefficients were used as features with a polynomial regression-based classification. A genuine acceptance rate of 96.08% was reported by using a data capturing device that comprised of 56 electrodes with a sampling frequency of 200 Hz. One major limitation of this database was that still only a single data recording session of 1 minute was made and thus it was not possible to assess the usability of EEG signals for biometric recognition over time.

AR-based features have also been tested using low-cost sensor system. Dan *et al*. [58] collected data from 13 healthy subjects using the NeuroSky MindSet single sensor headset. Data from three sessions were captured on different days; subjects were in a resting state with eyes closed during the data collection. A recognition rate of 87% was achieved using an LDA classifier. Comparing the results of [46] which used the same sensor device, it seems the AR features may be better suited in an identification scenario, while the PSD-based features may be a better choice for verification.

2.2.3 Wavelet Transform (WT)

Another relatively new and increasingly popular feature extraction method used for EEG biometrics is the Wavelet Transform. Wavelet-based features usually are derived from the wavelet coefficients $WT_{\psi}\{x\}(a,b)$, which can be computed using the following formula:

$$WT_{\psi}\{x\}(a,b) = \langle x, \psi_{a,b} \rangle = \int_{R} x(t) \cdot \psi_{a,b}(t) dt \tag{2.4}$$

where $x(t)$ is the time domain signal and $\psi_{a,b}(t)$ is the wavelet function. One advantage of wavelet-based methods is the flexibility of choosing the wavelet function $\psi_{a,b}(t)$*:* different wavelet functions with different scale (a) and shift (b) could be selected to suit specific applications. Moreover, the wavelet transform can

be designed to preserve signal content from specific ranges of both time and frequency in the wavelet domain which could potentially provide more relevant information in the feature space [59].

Gupta *et al*. [60] proposed the use of Wavelet Packet Decomposition (WPD) to extract three typical EEG bands (delta, theta and gamma bands) with Daubechies (db4) and Coiflet (coif3) wavelets for biometric identification. The EEG signals were triggered by visual stimuli and the P300 VEP data from four subjects was captured using eight electrodes (around 200 seconds of recording per subject) and used for feature extraction. An identification performance of 85% was reported using a Radial Based Function (RBF) Neural Network (NN) classifier.

Abdullah *et al*. [61] also explored the feasibility of using WPD for feature extraction. Their data were collected from 10 subjects, in 5 separate recording sessions over a course of 2 weeks. Each session consisted of 5 trials where each trial consisted of 5 tasks: eyes open, eyes closed, imagining right index finger movement, imagining left leg movement and puzzle solving. The mean, standard deviation, and entropy values of each wavelet coefficient vector were computed and used as features. 90% of all the data per subject were used for training a Neural Network (NN) classifier and the remaining 10% were used for testing: using just 4 electrodes (from the eight available electrodes) gave classification rates between 96% and 97%; using 2 electrodes gave classification rates from 90% to 95%. Using just a single electrode the results ranged from 70% to 87%.

2.2.4 Hilbert-Huang Transform (HHT)

The Hilbert-Huang Transform is a relatively new algorithm initially reported in 1998 [62]. Though it was designed for dealing with non-stationary signals, its use for EEG biometrics is still fairly rare. The HHT comprises two main steps: 1) in a process called Empirical Mode Decomposition (EMD) to generate the Intrinsic Mode Functions (IMF) and 2) performing the Hilbert Transform on each IMFs.

Given a signal $x(t)$, the effective algorithm of EMD can be summarized as follows [62][63]:

1) Identify all extrema of $x(t)$.
2) Interpolate between minima (*resp*. maxima), ending up with some "envelope" $e_{min}(t)(resp.\ e_{max}(t))$.
3) Compute the average $m(t) = (\ e_{min}(t) + e_{max}(t))/2$.
4) Extract the detail $d(t) = x(t) - m(t)$.
5) Iterate on the residual $m(t)$.

In practice, the above procedure has to be refined by a "sifting process" which amounts to first iterating steps 1) - 4) upon the detail signal, until this latter approach a zero-mean according to some stopping criterion. Once this is achieved, the "detail" is considered as the effective IMF, the corresponding residual is computed and step 5) applies [63].

The next step of HHT is for each IMF computing the Hilbert Transform. Denote the resulting IMF(s) after EMD by $x(t)$, the HT of $x(t)$ can be computed using (2.5):

$$y(t) = \frac{1}{\pi} P \int_{\tau}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{2.5}$$

where $y(t)$ is the result of HT and

$$P = \lim_{\varepsilon \to 0^{+}} \left[\int_{\tau}^{t-\varepsilon} \frac{x(\tau)}{t-\tau} d\tau + \int_{t-\varepsilon}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \right] \tag{2.6}$$

P is the Cauchy Principal Value defined by (2.6), due to the otherwise ill-defined function (2.5) while $\tau \to t$.

Kumari *et al*. [64] recently suggested using an EMD-based coefficient of variation (a parameter that relates to the standard deviation and mean of IMFs) as the feature for EEG biometric recognition. In a preliminary investigation, using data from only three subjects, they claimed that person identification performance is sensitive to the scalp regions from which the EEG signals are extracted.

2.2.5 Other Methods

Besides the previously mentioned feature extraction approaches which are based on the three major transforms, other interesting features have been developed for EEG biometrics. Singhal *et al*. [65] proposed a novel time domain peak matching

algorithm using VEP stimulated EEG signals for identification. They reported a recognition rate of 78% for a ten subjects' database. From each subject 20 VEP recordings (6 seconds each) were obtained with half of them used to train the system and the rest for testing.

Huang *et al*. [66] reported a simple but effective feature, tested using a relatively large VEP database (122 subjects). The proposed feature was the "equivalent root mean square (rms) values for each electrode signal over a 1 second period". This seems to capture the transient energy after each visual stimulus (1 second after the stimulus). As a time domain feature, its advantage was the simplicity of computation combined with a relatively high performance. The CRR reached 95.1% using the data from 64 electrodes for 116 subjects.

Event Related Potential was used for biometric recognition by Yearn *et al*. [67]. They claimed that showing the subject with self-face and non-self-face would trigger different VEP signals in the time-domain, which may be used for person recognition. A CRR of 85.5% was reported using a self-collected database which contained 10 subjects. The data capture device was the BCI2000 [19] and only 18 electrodes were selected for data collection.

Phung *et al*. [68] used the Shannon Entropy (SE) of alpha, beta, and gamma wave bands as features. The performance was compared with conventional AR-based features and achieved similar CRRs (97.1% versus 97.2%, for SE and AR). The advantage of the SE feature was claimed to be a much faster identification speed (2.3 to 2.6 times faster than using the AR features).

Gui *et al*. [45] employed two methods for EEG feature extraction: Euclidean Distance (ED) and Dynamic Time Warping (DTW). These two techniques were employed to measure the similarity between two EEG signal series for biometric recognition. The proposed methods were tested in a self-collected database with 30 subjects using a 74-channel EEG cap. Only the data obtained from four electrodes (Pz, O1, O2, Oz) were used. The ED method was reported to have over 80% identification accuracy whereas the accuracy of the DTW method was about 68%.

Conventional Mel-frequency Cepstral Coefficients (MFCC), normally used for voice/speaker recognition, were used as features for EEG-based person identification by Nguyen *et al*. [69]. For a population of 20 subjects (subset of a public database with 122 subjects) an identification rate of 92.8% was reported using the data from selected eight electrodes.

Considering these reported research trends, most of the feature extraction methods were based on visual stimuli. The experimental results using several newer features, despite their novel designs, were either comparable with conventional features such as AR coefficients and PSD ([66][68]) or worse than them ([65][67]).There is, therefore, a need for continued efforts to discover and develop new and more effective features for more robust EEG-based biometric recognition.

Wavelet-based and HHT-based features have shown significant potential in EEG biometric recognition, and it is, thus, interesting to develop further new features based on these transforms and evaluate them in larger databases (at least more than 100 subjects) extending over longer periods of time to also assess their template ageing performance.

2.3 Feature Classification

After the choice of features, the next most important component for an EEG biometric system is the design of the classification scheme adopted and strategy used to prepare the classifier for a particular application. In this section an extensive review of the existing works using different feature classification techniques for EEG biometrics is provided. Each subsection begins with a brief description of the classification technique, followed by its related literature survey for EEG biometrics: the literature review is focused on highlighting the techniques used for classification.

2.3.1 *k*-Nearest Neighbour Algorithms

One frequently used algorithm in EEG-based pattern classification is the *k*-Nearest Neighbour (*k*-NN) rule. The basic principle of *k*-NN algorithms for classification is to compare the similarity or distance between the template feature samples and the

query (or test) samples; the test set tends to find its *k* nearest labelled samples in the training feature space and the decision can be made by a majority voting scheme within the considered training samples: if the *k* is set to 1, the algorithm becomes a simple nearest neighbour classifier and the decision making is solely dictated by the label of that nearest neighbour's class [70].

The learning process of *k*-NN directly uses the available training set, which makes it an instance-based machine learning algorithm [71]. Therefore, the classification is sensitive to the local (geometrical) distribution of the training samples, which may potentially lead to the instability of performance [72]. For instance, with the training set increasing, the local training feature distribution near a test set may change as well, which may lead to a change of the membership of the nearest neighbours and the final decision when *k*-NN is applied. Another shortcoming of the basic *k*-NN algorithm is that the density of the data clusters affect its performance and may lead to wrong decision making: depending on the parameter *k*, the decision may be biased towards the class with a high density of the cluster, and makes the *k*-NN classifier sensitive to noisy data [70]. To overcome this potential problem, some *k*-NN algorithms add the distance (*d*) weight (conventionally 1/*d*) as a contribution for decision making as well [70].

Despite these shortcomings, *k*-NN does have one significant advantage: it does not have an explicit and distinctive training process. The available training samples (multi-dimensional feature vectors) only need to be stored together with their respective class labels and compared with the query samples directly. Such simplicity is a clear advantage over many other machine learning algorithms. Some of the following reports highlight this advantage (fast classification process) while employing the *k*-NN classifier for EEG biometric recognition.

Palaniappan and Ravi [73] reported using a database which contains 20 subjects for identification. Their sensor system comprised of 61 electrodes, all of which were taken into consideration. The gamma band PSD feature was extracted and the Principal Component Analysis (PCA) was used for noise removal. Three different classifiers were employed for testing the classification performance: Simplified Fuzzy

ARTMAP (SFA), linear discriminant (LD) classifier and the *k*-NN classifier. Euclidean and Manhattan distances were used to locate the nearest neighbour for *k*-NN. The reported CRRs of LD and the *k*-NN were almost the same (both around 92%), though the highest performance was obtained by using LD (96%), but using *k*-NN for the system training was found to be 3 times faster than SFA and 1.5 times faster than LD.

Yazdani *et al*. also employed the *k*-NN classifier for person identification using VEP data [74]. A 61-electrode sensor system captured the EEG data from 20 subjects. AR and PSD coefficients were used as features. They claimed a 100% accuracy rate using a 5-NN classifier: the parameter *k* for *k*-NN was extensively tested from 1 to 100. In these experiments the optimal number of *k* in EEG biometrics usually tends to be relatively small.

Fei *et al*. used data captured with a single electrode (Fp1) from 40 subjects [75]. An identification accuracy of 97.5% was obtained using a *k*-NN classifier along with Fisher's linear discriminate analysis for feature reduction. The recognition performance achieved using a support vector machine classifier was 81% in their experiments.

2.3.2 Linear Discriminate Analysis

According to its frequent appearance in the relevant literature and its high recognition performance, linear discriminate analysis (LDA) is by far one of the most popular classifiers in EEG-based biometric classification, For a classification problem, LDA assumes that the conditional probability density functions of multiple classes are normally distributed with equal class covariance[76]. In the Bayesian framework, the optimal solution can then be approached by computing the ratio of the inter-class variance σ^2_{inter} over the intra-class variance σ^2_{intra}:

$$S = \frac{\sigma^2_{inter}}{\sigma^2_{intra}} \tag{2.7}$$

The score *S* of (2.7) is the likelihood ratio, which can be viewed as a threshold (often preferred to be maximized) in classification problems [77].

A closely related technique to LDA is the so-called Fisher's LDA, which also occasionally appears in EEG biometrics literatures. The only difference between Fisher's LDA and the conventional LDA is that the Fisher's LDA does not make some of the assumptions of LDA such as normally distributed classes or equal class covariance [77].

The work in [78] is one of the early reports which employed the LDA classifier to investigate the effectiveness of EEG signals for biometric person recognition. Publicly available data for five subjects performing mental tasks were used. A series of different features were computed, including AR coefficients, channel spectral powers, inter-hemispheric channel spectral power differences and inter-hemispheric channel linear complexity. A standard PCA algorithm was employed for feature dimension reduction and the resulting features were fed to a LDA classifier. The reported results indicated a minimum average recognition error of 0%.

Sharing certain similarity with PCA, LDA could also be used for feature dimension reduction [74]. One of the arguable advantages of LDA (and particularly Fisher's LDA) over PCA is that LDA takes class differences into account when performing dimension reduction [79]. In [74], Fisher's LDA was used for dimension reduction before the feature vectors were fed into an optimized *k*-NN classifier (5-NN). Data from 20 subjects, contained in a publicly available database (with 122 subjects), was used for their experimental tests. All the data obtained from 61 electrodes were used for feature extraction: conventional PSD and AR coefficients were used as features and fed to a *k*-NN classifier and an accuracy of 100% reported.

In [80] an LDA classifier was used with the spectral power, the maximum power, and the frequency of maximum power in the alpha band as features. Two sessions of data from four subjects were collected using "one bipolar channel (O1A2)", and the time interval between the sessions ranged from 10 days to 5 months. The "authentication performance" was reported as 98.33% with test recordings of 20 seconds duration. The results suggest robustness to template ageing (as opposed to the report in [81], which shown 10.9% of identification rate in the repeatability investigation).

2.3.3 Artificial Neural Networks

Artificial Neural Network (ANN) forms a family of statistical learning algorithms inspired by biological neural networks (in particular the brain). These can be used to estimate or approximate functions that may depend on a large number of generally unknown inputs [70]. In general an ANN can be viewed as a network of simple classifiers; each classifier in the network is an activation function which only responds to the input from its previous neuron (function). The results of these functions are weighted and fused before they reach the final decision making.

A typical ANN is defined by the following parameters [82]:

- The interconnection pattern between the different layers of neurons
- The learning process for updating the weights of the interconnections
- The activation function that converts a neuron's weighted input to its output

ANNs have been popular classifiers for EEG biometrics since this modality was first explored [6]. The work reported in [83] is one of the early studies which employed this kind of classifier for EEG-based person identification. In that study a "multilayer perceptron NN with a single hidden layer" was used for classifying the data obtained from 20 subjects. A headset with 61 electrodes was used to capture 40 seconds of VEP data (in a single recording session). An identification accuracy of 99.06% was reported by computing the conventional PSD-based features from gamma band data (32Hz-40Hz).

The report in [84] further increased the test population to 40 subjects, with the experimental framework following the work in [36]. Using the same feature and headset as in [83], the correct recognition rate increased from 99.06% to 99.62%. Half of the available pattern samples (i.e. 20 seconds from each subject) were used for training while the rest were used for testing. A three-layer Elman Neural Network (ENN) was used for classification. A detailed rationale for the design of the ENN may be found in [85].

Using a single electrode (Cz) in an ANN-based identification scenario was investigated in [86]. Data from only three subjects was obtained was filtered to retain

the frequency ranges from 1 Hz to 12 Hz. The reconstructed time domain data was fed directly to a single hidden layer ANN, trained by the data from five trials (1 second of visual stimulus per trial), and an accuracy of 100% was reported.

Using a low-cost sensor system for VEP data collection was investigated by Gui *et al*. [87]. They reported of using a 6 midline electrode system ("EASY CAP"[88]) in an identification scenario for 32 subjects. Only the data obtained by the electrode placed in the Oz location was employed for feature extraction. The WPD was applied for noise-removal; the mean, SD and entropy of the resulting wavelet coefficient were computed as features. The VEP data was collected during the subjects reading. Each recording lasted for 1.1 second, the overall recording length contained about 55 seconds per subject. 70% of the obtained data was used to train a "feed-forward, back-propagation, multi-layer perceptron" neural network and the rest of the data was used for testing the recognition accuracy. A performance of around 90% in a one-against-the-rest identification scenario (binary classification) was reported. They also reported that, in a conventional identification scenario (32 classes) the obtained accuracy was less than 11%, which indicated the low-cost single sensor system was still not reliable enough for realistic biometric scenarios.

The quantity of data and time required for training ANNs is a major concern for effective deployment. The reported researches on EEG biometrics usually employ ANNs with simple structures (e.g. a single hidden layer). With the improvement of computational power, i.e. faster central processing units (CPU), and the employment of graphics processing units (GPU), the use of more complex networks, such as recurrent and convolutional neural networks, have shown great potential in pattern recognition [89]. The concept of "deep learning" has been quite successfully implemented in many pattern recognition fields, including handwriting recognition and fast image processing [90]. Such classifiers hold great promise for application in EEG-based biometrics recognition.

2.3.4 Kernel Methods

Kernel-based classifiers were proposed as early as the 1960s, with the invention of the kernel perceptron [90][91]. These algorithms became prominent with the popularity of the support vector machine (SVM) in the 1990s, when the SVM was found to be competitive with neural networks on tasks such as handwriting recognition [92].

The SVM classifier integrates the notion of supporting vectors and kernel tricks. Given a set of training samples, SVM is designed to find the "maximum-margin hyper plane" in feature space which indicates the maximum gap between two classes: the feature vector(s) which form the hyper plane(s) are called the "supporting vectors" [92]. However, it often happens that the available data are not linearly separable in the original feature space. Therefore, it was proposed by Vapnik *et al*. [92] that the original finite-dimensional space may be mapped into a space with much higher-dimensionality to make the separation easier in that space. The mapping of the original space onto higher/infinite spaces is called the "kernel trick" [91]. Some commonly used functions (kernels) for this mapping in SVM are polynomial functions, the Gaussian radial basis function, and the hyperbolic tangent function [92].

The classic SVM is a binary classifier and some combination strategies have been proposed to make it applicable for multi-class model training (more than two classes). Two conventional strategies are [93]: building binary classifiers which distinguish (*i*) between one of the labels and the rest (one-versus-all) or (*ii*) between every pair of classes (one-versus-one). Classification of new instances using the one-versus-all strategy is followed by a winner-takes-all rule, in which the classifier with the highest output function assigns the class. It is important that the output functions should be normalized to produce comparable scores. For the one-versus-one approach, classification is done by a max-wins voting rule, in which every classifier assigns the instance to one of the two classes, then the vote for the assigned class is increased by each new vote, and finally the class with the most votes determines the instance classification [94].

In the field of EEG biometrics, the use of SVM classifiers has also received attention. Using conventional AR coefficients features for classification, results were presented using three classifiers, namely Linear Discriminate Analysis (LDA), back-propagation Neural Network and Support Vector Machine in [95]. The reported accuracies, using a small publicly available database with 3 subjects and using five electrodes, indicate that ANN and SVM provided comparable performance (identification accuracy ranges from 80.8% to 84.0%) and LDA outperformed both of them by more than 5% (89.5%).

The SVM classifier was used with the data obtained using a low-cost sensor system (Emotiv [20] with 14 electrodes) in [96] for 5 subjects. Each of them "performed four different mental tasks" with an overall recording length of 150 seconds. A series of different features were extracted: 6th order autoregressive (AR) coefficients, power spectral density, and total power in five frequency bands, inter-hemispheric power differences and inter-hemispheric linear complexity. These features were fed into a one-versus-all linear Support Vector Machine [94]: for every subject, the data of four segments from four tasks (10 seconds per tasks) were combined and used for testing the performance, the rest of the data was used for training the SVM. Classification accuracy of 100% was reported through the combination of two voting rules by computing the one minus the averaged half total error rate.

To summarize the characteristics of the classifiers reviewed in this section, (of the four mentioned mainstream classification algorithms in EEG biometrics) those based on the *k*-NN algorithm tend to require the least computation time for model training, while maintaining acceptable recognition performance. Compared to *k*-NN classifiers, the SVM and ANN classifiers are more complex and they have the potential to achieve much better performance in EEG biometric recognition. Especially with the recent development of "deep learning" algorithms, it is worth investigating the application of improved ANN techniques to EEG biometric recognition. The challenge here would be the availability of substantial training data would be required to train such large networks. The LDA algorithm has demonstrated very good performance amongst the results reviewed, and unlike the SVM and ANN classifiers

which can be quite time-consuming in training, LDA has a satisfactory trade-off between training/testing time and classification performance.

2.4 Usability Investigation of the Reported Results

In this section, five factors affecting usability in practice are investigated and combined to provide one single metric to provide an indication of the potential usability of the EEG-based biometric systems. The use of EEG data for person recognition has been researched since 1998 and more than 100 papers in this field have been published. Most of these papers focused on the performance using conventional accuracy metrics (in both the identification and verification scenarios). However, the EEG signal as a biometric modality contains its particular characteristics: to objectively evaluate an EEG biometric recognition system and assess its potential suitability for real-life applications, the performance of EEG-based biometrics system needs to be evaluated based on more factors than the recognition accuracy only.

Besides the conventional accuracies, there are at least another four factors which should be taken into consideration for assessing the practical usability of the reported EEG-based recognition systems: 1) number of the subjects for which the system was designed or tested on (N); 2) number of the electrodes employed (K); 3) the recording duration of the training set (Tr) and 4) recording duration of the test set (Te).

The number of electrode(s) used for the data collection has a large impact on the usability of the EEG-based biometric system: small number of electrodes is expected in the real-life scenarios. The EEG recordings employed for system training, and particularly the time required to achieve the person recognition are quite influential factors. The number of the subjects employed also indicates the capability of a given biometric system. Training and test durations (in second) are proposed to be used here instead of the number of samples, as in the actual application it is the time to be spent by user(s) and the computational efficiency that matters.

Including the correct recognition rate (CRR), five factors in total may be considered together in evaluating a system. Of these five factors, K, Tr and Te are should ideally

be as small as possible for practical deployment; N and CRR should be as large as possible. Therefore, the following metric (2.8) is proposed to combine these factors and produce an overall score, U, to evaluate the system:

$$U = \frac{N \times CRR \times 100}{Tr + K \times Te} \tag{2.8}$$

This score which indicates the "effectiveness" or usability of an EEG biometric system can therefore be used to provide a meaningful comparison of the systems that have been covered in this review: the higher the value of U, the better the system's overall usability. The time required for the user to provide data in normal use is represented by the testing time (Te) and is considered to have a greater weight in establishing the usability of the system compared with the one-off time required for training Tr: since the tests may be conduct multiple times, the number of electrodes K is used as the weighting parameter in (2.8).

In the following subsections the proposed metric is used to compare the proposed systems covered in the review. These are grouped by the different feature types that have been used for classification, as it was shown in Section 2.2 that the choice of feature is most crucial in developing EEG-based biometric system.

2.4.1 PSD-based Systems

As it was stated in Section 2.2.1, PSD features have been amongst the most popular used for EEG biometric recognition. Table 2.2 compares, with the help of the proposed metric, the effective usability of some reported systems that employ these features. The order of the listed papers follows the year of the publication in ascending order. As it can be observed in Table 2.2, the overall usability scores (U) generally tend to increase as the year of publication increase, which indicating a trend of improving usability of EEG biometric systems. The highest score listed in the table reached S=33.33, which indicates the best overall performance using PSD-based features amongst the papers covered in this review.

Table 2.2 Identification Comparison of the reports using PSD for feature extraction

Reports	*N*	*K*	*Tr(s)*	*Te(s)*	*CRR(%)*	*U*
Poulos *et al*.99'[5]	4	1	3600	4500	95	0.05
Palaniappan *et al*.02'[36]	10	61	20	20	95	0.76
Palaniappan 03'[38]	20	61	20	20	94.18	1.52
Palaniappan 04'[83]	20	61	20	20	99.06	1.59
Ravi.*et al*.05'[39]	40	61	20	20	96.63	3.13
Palaniappan *et al*.07'[41]	40	35	20	20	98.56	5.56
Palaniappan *et al*.07'[42]	102	61	20	20	98.12	8.33
Sun Shiliang 08'[97]	9	15	540	270	95	0.1
F.Su *et al*.10'[30]	40	1	1800	1800	97.5	1.09
Hema *et al*.10'[98]	15	2	160	40	89.95	5.56
Zhao *et al*.10'[99]	10	1	100	50	96.77	6.25
Quintela *et al*.10'[100]	70	8	45	45	99.1	8.33
Rocca *et al*.14'[31]	108	64	50	10	100	16.67
Bai *et al*.14'[101]	20	32	10.8	1.2	97.25	**33.33**

2.4.2 AR-based Systems

Feature extraction based on AR coefficients is another popular approach in EEG biometrics (Section 2.2.2); some relevant reports with experimental details are compared in Table 2.3 using the proposed usability metric. Based on this metric, the best performed system using the AR features performed better than the best system using the PSD-based features amongst the papers reviewed in this survey.

Table 2.3 Identification Comparison of the reports using AR coefficients for feature extraction

Reports	*N*	*K*	*Tr(s)*	*Te(s)*	*CRR(%)*	*U*
Poulos *et al*.99'[102]	4	1	3600	4500	72~84	0.04
Poulos *et al*.99'[5]	4	1	1800	6300	91~97	0.05
Paranjape *et al*.01'[54]	40	1	34	34	80	**50**
Poulos *et al*.02'[103]	4	1	3600	4500	99.5	0.05
Palaniappan *et al*.05'[43]	4	6	50	50	99.04	1.14
Palaniappan *et al*.06'[78]	5	6	50	50	99.88	1.43
Yazdani *et al*.08'[74]	20	61	89	1	100	0.36
F.Su *et al*.10'[30]	40	1	1800	1800	97.5	1.09
Zhao *et al*.10'[99]	10	1	100	50	96.77	6.25
Campisi *et al*.11'[57]	48	3	43	17	96.08	**50**
Kostilek 12'[81]	9	53	45	15	98	1.05
Rocca *et al*.12'[104]	45	3	43	17	98.73	**50**
Dan *et al*.13'[58]	13	1	9360	4680	87	0.08

2.4.3 Other Feature-based Systems

In this section papers reporting other features for use in EEG biometrics are compared using the proposed metric in Table 2.4. The usability score *U*=100 is achieved using a wavelet-based feature, which is by far the best-reported performance amongst the EEG biometrics papers covered in this survey.

Table 2.4 Identification Comparison of the reports using other algorithms for feature extraction

Reports	*N*	*K*	*Tr (s)*	*Te(s)*	*CRR(%)*	*U*
Gupta *et al*.09'[60]	4	8	132	68	85	0.51
Yang *et al*.12' [105]	18	8	240	60	97.4	2.43
Yeom *et al*.13'[67]	10	18	360	40	85.5	0.79
Yang *et al*.13'[106]	109	8	240	60	90	14.29
Rocca *et al*.13'[107]	36	3	40	20	99.69	33.33
Yang *et al*.13'[108]	50	5	114	6	95.5	33.33
Phung *et al*.14'[68]	40	23	800	400	97.1	0.39
Yang *et al*.14'[7]	105	1	96	24	99	**100**

A cumulative illustration of the reported EEG biometric systems is shown by Fig. 2.3. The usability scores of the reported systems are summed per year. It is noticed the score has a sudden increase in the year 2001, the possible reason of such a boost may have to do with a manually data cleaning process in that experiment: “the muscle (EMG), cardiac (ECG), or other noise signals were removed” by “a trained neurologist” [54]. The usability of EEG biometric received consistent increasing in recent five year, reflected from both the growing number of publications and the better usability system design.

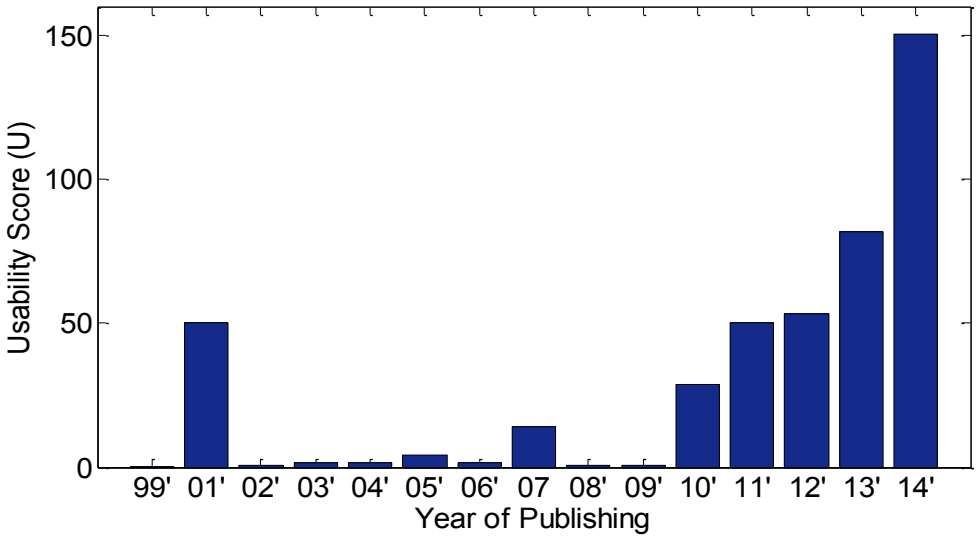


Fig. 2.3 Yearly plot of the usability scores

In summary, the usability scores of the reported systems seem to indicate some of the early reported EEG biometric systems may be focusing only the recognition performance and ignored other important factors, such as the number of electrodes and the recording length. The proposed usability metric reveals the gap between the laboratory context where these systems are developed and the real-world environments of target applications for EEG biometric recognition. However, due to the limited number of the available reports related to template ageing, the proposed metric for computing the usability score has not considered the template ageing effect in EEG biometric recognition.

In biometric person recognition, template ageing is an important consideration that may limit the usability of biometric systems in terms of the need for frequent acquisition of new enrolment data. The comparisons reported in these tables were either obtained using single session EEG data or trained using concatenated data from multiple sessions. Therefore, the factor of template ageing has been excluded from the proposed metric.

2.4.4 Template Ageing

Besides the system training and test durations, other EEG-based characteristics also impact the usability this modality considerably. Template aging is one such important

factor which needs to be considered before EEG biometrics becomes widely applicable.

Lee *et al*. [80] investigated the separation of data from multiple sessions for the system training. EEG was recorded while subjects were resting with eyes closed during two sessions separated by time intervals ranging from 10 days to 5 months: data of the first session was used as training and the second session's recording was used for testing. An accuracy of 98.33% was reported with 20 seconds of recording for testing (20s of training and 20s of testing). However, the database contains only four subjects.

Visual-ERP data captured from a single sensor at the Oz position was used for template ageing analysis for biometric recognition by Blondet *et al*. [109]. They conducted the experiments with 15 subjects for multiple sessions with a time interval between the first two sessions ranging from 5 to 40 days. Out of those participants, nine returned for a third time, after 134 to 188 days from the first session (mean: 156 days or 5 months). The time domain cross-correlation was computed for feature extraction. An average accuracy of 0.99 was reported for the first session, 0.93 for the second session and 0.84 for the third session".

Template ageing effects when using mental-tasks with long time interval between training and testing has also been noticed. In [81] , Kostilek conducted experiments using 53 electrodes and 9 subjects (using actual and imagined motor tasks) from two sessions, with a time interval of approximately one year. Using part of the data from the first session for training and the rest of the first session's data as the test set, the CRR reached as high as 98%; whereas using the first session's data for training and data of the second session for testing, the performance reduced to a CRR of 87.1%.

These results suggest that while template ageing may result in loss of accuracy in EEG-based biometric systems, the impact of ageing may not be catastrophic. However, the results reported so far are for very few subjects and relatively short temporal separations. More research is required to quantify the nature and extent of template ageing effects in EEG biometrics.

2.5 Conclusion

This chapter provides a review of the literature on EEG-based biometric person recognition. As it is a relatively new biometric modality, the literature is focused on establishing the presence of biometric information in EEG signals. The use of EEG biometrics in real-world application scenario would need much more research to address the shortcomings of the work done to date especially with regards to the quantity and nature of data available for system evaluations. A novel usability metric was developed and evaluated using the existing literatures; the proposed metric better represented the usability characteristics of the state-of-art EEG biometric systems.

The search for better acquisition systems, signal preparation, feature extraction and classification should target the goal of reducing the sensor costs (electrode numbers) and quantity of data needed for training testing recognition systems while maintaining high recognition rates. Quantifying and ameliorating the effects of template in EEG biometrics is also an important area for further research.

Chapter 3

Experimental Framework

The overall experimental framework included in the thesis is presented in this chapter. Two publicly available databases and one self-collected database employed in the experiments are described in Section 3.1. Evaluation methods are introduced in Section 3.2, which are used throughout the work involved in the thesis. The conclusion is presented in Section 3.3.

3.1 Databases

The existence of some physical and psychological differences between brains should be without doubt; however, as unique as the brain is, whether the EEG signals, which obtained by sensors with different level of sensitivity still preserve enough information that individually distinctive may be debatable. Using EEG signals for biometric purposes is still a relatively new direction for research (15 years since the first relevant paper was published) and there has been no publicly available EEG database which especially collected for biometric purposes.

Three databases were used in the research: two publicly available databases which were collected using a sophisticate multi-sensor system; and one self-collected database using a low-cost single sensor headset which caters to the requirements of biometric recognition scenarios. The following sections provide details of these databases and their particular characteristics, which may reflect onto the subsequent experimental analysis in the following chapters.

3.1.1 PhysioNet EEG Motor Movement/Imagery Dataset

EEG Motor Movement/Imagery Dataset (MM/I dataset) contains the data obtained from 109 subjects, provided by developers of the BCI2000 instrumentation system: the headset contains 64 wet sensors with sampling frequency of 160 Hz (10-10

electrode positioning system [18][110]. The electrodes positioning map is illustrated in Figure 3.1.

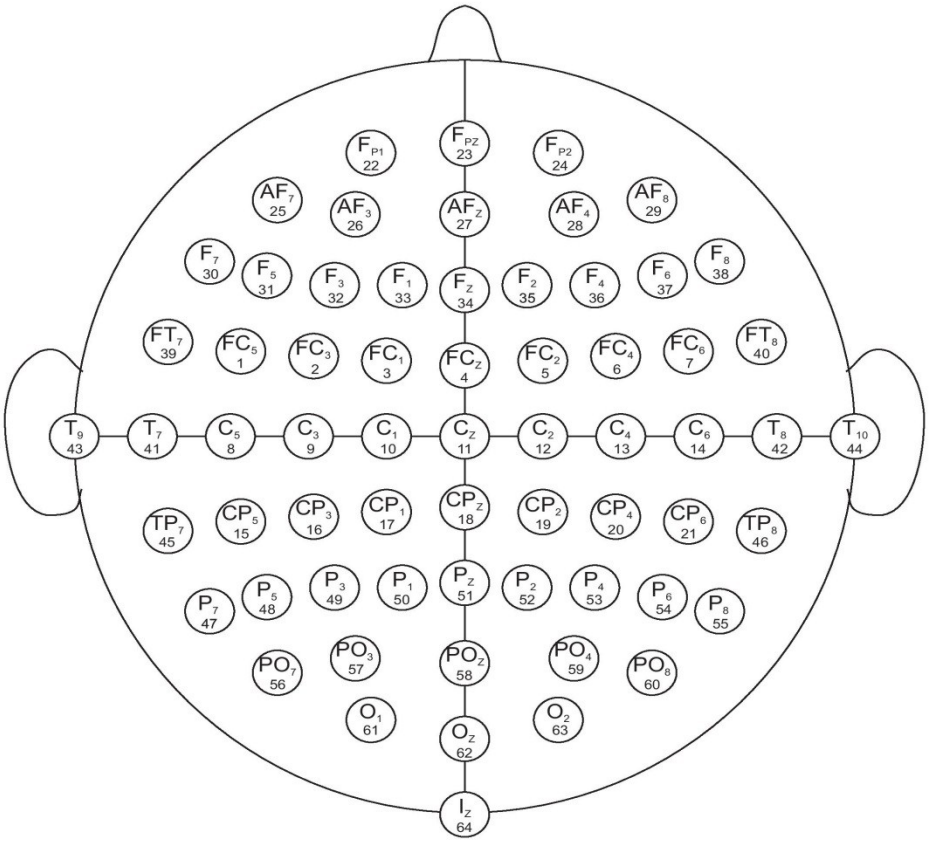


Fig. 3.1 64 electrodes positioning based on 10-10 system

Subjects involved in their data collection performed six different tasks: two baseline tasks where there is no requirement for a specific activity (one with eyes open, one with eyes closed) and four motor movement/imagery tasks. Details of these four tasks are detailed as follows [15]:

a) A target appears on either the left or the right side of the screen. The subject opens and closes the corresponding fist until the target disappears. Then the subject relaxes.

b) A target appears on either the left or the right side of the screen. The subject imagines opening and closing the corresponding fist until the target disappears. Then the subject relaxes.

c) A target appears on either the top or the bottom of the screen. The subject opens and closes either fists (if the target is on top) or feet (if the target is on the bottom) until the target disappears. Then the subject relaxes.

d) A target appears on either the top or the bottom of the screen. The subject imagines opening and closing either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. Then the subject relaxes.

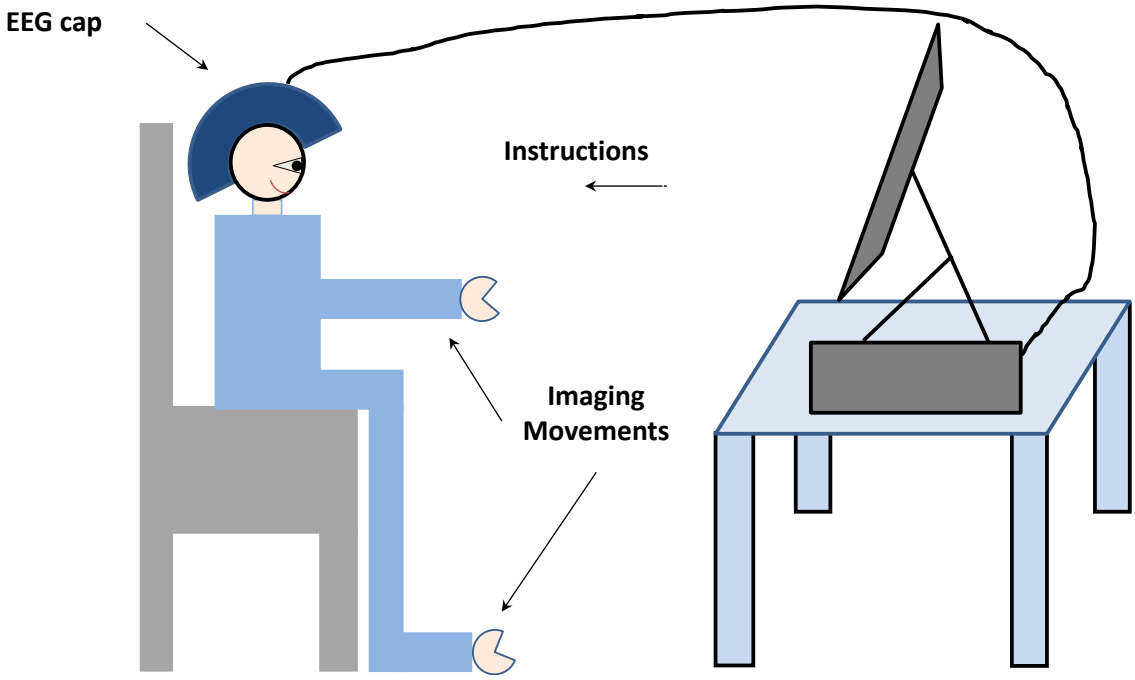


Fig. 3.2 Data capturing set up diagram

The diagram of the data collection set-up is shown in Fig. 3.2. Three runs were performed for each subject per motor task, and each run lasted for two minutes. The baseline tasks were performed only once and lasted for only one minute. As the time interval of each run was a couple of minutes (according to the contributor of the database), it is assumed the EEG headset was not removed between each recording run. Since the actual recording lengths for each subject vary (data of some subjects contain much less than 2 minutes' recording per run and some contain more than 2 minutes), for some of the experiments included in the thesis, only data of the selected

subjects were employed to guarantee the data with equal recording length for system training.

As one of the pioneering researchers who adopted this database for biometric experiments (if not the earliest), the author of this thesis believes MM/I dataset indeed contains certain characteristics which may be of helpful in simulating the possible real-life EEG biometric scenarios. For example, the available recording length of MM/I database per run is relatively short (each run 2 minutes), which may be an acceptable time length for the system training. Despite the large number of electrodes used during data collection, it is possible to use a subset of these (even data from one single sensor) to simulate a more realistic biometric scenario. Additionally, the MM/I database contains multiple runs of recording, though strictly it cannot be considered as a database with true multiple sessions for template ageing effect analysis, the database is still good enough to explore the short term stability of EEG signals as carriers of biometric information.

The experimental results seem to suggest that the quality of the MM/I database is the highest amongst the three databases included in the thesis: depending on the particular method/algorithm used for biometric recognition, the noise removal may not even be necessary to obtain satisfactory results. However, as it was mentioned, the MM/I database in fact does not contain the data from multiple sessions that separated in long time interval, the results obtained from multiple runs thus may not be representative for the real-life scenarios as the template ageing/consistency effects must be considered.

3.1.2 UCI EEG Database Data Set

The "UCI EEG Database Data Set" [111] with different stimulus was used to compare with the MM/I dataset. It contains data obtained from a comparable number of subjects (122 subjects) with the MM/I dataset, but the stimulus belongs to a different category: visual-based stimulus. Research conducted in this thesis investigates the impact of the different stimuli and evaluates the robustness of new algorithms. This database is also referred as VEP database henceforth in this thesis.

The VEP database was also collected by a sophisticated headset/sensor system with 64 electrodes (10-20 electrode positioning system). 122 subjects were separated into two groups: alcoholic (77) and control (45). The original purpose of the database was to investigate the impact of alcoholism on human brain. Depending on different individual, the data recording ranges from around 15 seconds to about 2 minutes in a single data collecting session. In some of the experiments reported in this thesis, only the selected 118 out of 122 subjects were employed for the same reason as the case of MM/I database (to guarantee equal length of recording for data analysis). The distinction between two types of subject (alcoholic and control) was ignored in the experiments and only the biometric performance explored in the thesis.

During the data collection, subjects were stimulated through viewing a series of standard picture sets (banana, airplane etc.) while their EEG signal was recorded [37]. Each picture was shown for one second and, the following one second's EEG data was recorded (with a sampling frequency 256 Hz) as one run. There were between 15 to 120 such one-second trials recorded, depend on different subjects. Therefore, the overall recording length varied a lot for different subjects. Further detail of the data capture and preparation of this database can be found in [37].

One potential advantage of VEP database over MM/I database is that, were the research focused on detecting a particular pattern which related to the stimulus, the VEP database may be better in revealing the corresponding patterns: visual stimuli are bound to trigger certain EEG waveforms (such as P300, for example [35]) for healthy subjects, however, motor-based EEG reflections in time domain may be hard to justify. One limitation of the VEP database however, is the fact that it also contains only a single session of data recording for each subject. Both of MM/I and VEP databases are good choices in justifying the EEG signals contain distinctive biometric information, but they lack the potential for investigating whether EEG signals are stable enough over a relatively long period of time. Furthermore, as the VEP patterns are similar between different individuals, the between-class similarity may cast negative effects while using VEP-triggered waveforms for biometric recognition.

3.1.3 Mobile Sensor Database

Due to the lack of suitable publicly available databases, there is a necessity to collect a database that especially caters to the evaluation of EEG signals' biometric potential in the realistic scenarios. A self-collected Mobile Sensor Database for this research was particularly designed to maximize the mimicking of real-life scenarios as much as possible:

1) Two sessions of data was separately collected, the time interval between sessions ranges from three weeks to two months;
2) The data was recorded by a low-cost single dry sensor system, which simplifies the deployment of the headset (system set up only requires several seconds' cooperation from user);
3) Subjects were asked to perform a simple mental number-counting activity with eyes closed.
4) The data used for training the biometric system in the experiments was one minute's long; each of the test duration was only around ten seconds.

The motivation of instructing subjects to perform a simple mental counting without much of restriction in content is based on some preliminary results obtained by analysing the MM/I dataset (see Chapter 5): using different motor tasks for system training and testing do not seem to be quite sensitive to biometric performance, as long as the training and test sets are the same type. This may suggest that better performance tend to be achieved while brain is relatively active; and the performance of biometric recognition may not be quite sensitive to particular stimulus. The Mobile Sensor Database comprises of multi-session data from 27 subjects collected using the NeuroSky MindWave headset [17] (Fp1 electrode). Seven of them are research lab members and have basic understanding of experimental data collection, while the other twenty volunteers are less professional and relatively careless during the data capture. The experimental scenario follows the scheme of using the whole one minute's data of one session for training and another 10 seconds' continuous data (randomly selected for a six-fold cross validation) from the other session for testing

(both of the two sessions lasted for one minute in all). Figure 3.4 illustrates the picture of MindWave EEG headset.

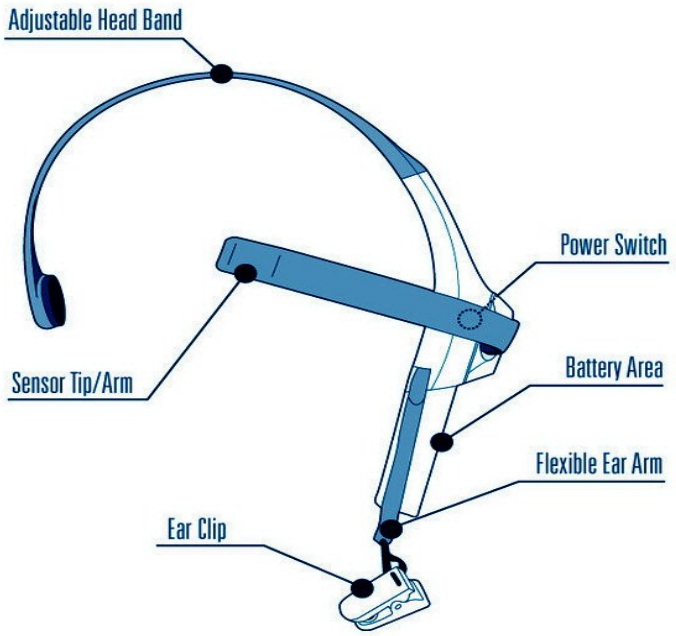


Fig. 3.3 NeuroSky MindWave Headset [6]

It is good to clarify one fact that distinguishes the EEG-based pattern recognition from some other biometric recognition tasks: unlike most of the physical traits (such as iris and fingerprint), or handwriting and speaker recognition which depend on conventional behavioural traits, EEG recognition lacks the possibility to confirm the quality of the ground truth data. For example, in most of image-based pattern recognition, without considering the performance of the recognition system, it is almost always possible to manually verify the quality of the image: if the image is good but the algorithm performs badly, it is certain that something has gone wrong with the method/algorithm; whereas in the case of EEG signals, if the performance is poor, there may be no simple way to check whether it is because the quality of the raw data is too low or the algorithm does not work (though if the signal were too noisy it can be visible).

The clear raw EEG signals are very similar with each other in shape, they look like a series of naturally encrypted codes: this may be its advantage as a biometric modality (no point of stealing the data if there is no appropriate method to decode it and match

it with template), but it also introduces a lot of difficulties in research. This perhaps is another reason why many published results so far prefer to use the VEP-based signals for biometric analysis: despite the P300 wave which triggered by the visual stimulus shares similar shape for every healthy individual, at least there is something that can be seen and distinguished.

Text-independent speaker recognition, on the other hand, does share certain similarity with EEG recognition: as opposed to being unable to verify the ground truth signals as the case of EEG (here the "ground truth" means the visually obvious characteristics of the signal, not the signal as a biometric identifier), the time domain detail of the signals are often ignored in text-independent speaker recognition, the language meaning which the speaker produced usually is not taken into consideration in the implementation. Furthermore, it is always possible to hear the sound of the speaker and check whether it is too noisy or not, which is something that may be more difficult for EEG signals. Theoretically, brainwave should be a thought-dependant modality, but due to the restrictions of the employed databases/headset mentioned previously, in this research some algorithms assume EEG signal (not brainwave) is a thought-independent modality.

3.2 Evaluation Methods

Typical biometric systems usually are implemented in two scenarios: identification and verification. In this thesis, both the scenarios are experimentally investigated using EEG signals. Identification is a one-to-many testing scenario: the identity of the query (test set) is not available a prior, the tests including multiple similarity comparisons between the query and all of the templates in the training set. Two possible ground truth cases may exist in identification scenario: 1) the identity of the query belongs to one of the templates in the database or 2) identity of the query is not enrolled in the system and no template stored in the database [2].

In the first case, the True Positive Identification Rate (TPIR) which measures the successful rate of correct matches is usually computed to evaluate the performance of the system. Performance of the second case may be described by False Positive

Identification Rate (FPIR), which measures the probability of a positive match between an un-enrolled query and one of the templates in the database. The FPIR depends on both the size of the enrolled database and a customized threshold to sift out the imposter query. The computing of FPIR is not often required in biometric scenario (as opposed to the forensic scenario) since it is overlapped with the verification scenario in its real-life applications [112] [113].

On the other hand, the TPIR is used as a standard metric to evaluate the biometric identification system: since the potential positive matching candidate (template) of any query set is guaranteed within the database, instead of customising the threshold in the case of computing FPIR, a process of ranking is performed to identify the best template for any particular query. This ranking provides a series of matching templates sorted by their respective probability scores, the performance derived from the highest score ($s = 1$) of each class is called the rank-one accuracy (often simplified as "accuracy" in the literatures). While $s = k$ ($1 < k < N$, N is the number of class), the matching allows to be less restrictive and the so-called Cumulative Match Characteristic (CMC) curve may be generated to illustrate the sensitivity of the identification system to the test set(s) [114]. The identification performances of EEG biometric systems report the thesis were evaluated using both TPIR and CMC curve.

Verification is another important biometric scenario. Some recognition systems proposed in the thesis were evaluated in this scenario. Different from identification scenario, verification is a one-to-one testing process, i.e. the test set with a claimed an identity is fed into the system and only the template linked to the claimed label in the database is compared with that test set. Same as the identification, there are also two cases may occur in the scenario: 1) genuine case and 2) imposter case.

The genuine case is a relatively simple situation compared with the imposter case. If the real identity of the test set is what it is claimed, the system will return two feedbacks depend on the customising of the thresholds: the successful template matching or the template mismatching, they are conventionally measured by the False Rejection Rate (FRR) and False Acceptance Rate (FAR), respectively [115]. The probabilities of FRR and FAR have inversed trends, while the FRR reducing the FAR

tend to be increasing, but not necessarily strictly proportional. Depending on particular applications, the compromised point between FRR and FAR could be purposely tweaked/optimised.

The imposter case though is also measured using FRR and FAR, the source of the data may be more diverse. For experimental simulation, two schemes may be derived to evaluate this case: 1) using other subjects' query sets which have corresponding templates saved in the database as the imposters for testing. In this scheme, the data of the imposters are fixed and their volumes only depend on the available database itself. However, in scheme 2) it is also possible that the data source of the imposters beyond the available database: theoretically it can have infinite number of imposters to spoof the system. Based on this perspective, any biometric system will eventually fail given enough attempts of imposter attack. Therefore, the verification systems developed in the thesis only followed only the first scheme.

Due to the limited amount of available data for the experiments, cross-validation was employed to avoid the possible results bias [116]. Based on the rationale of this evaluation method, the available data is normally equally divided into k folds and the rotating estimation throughout all the folds is performed. The accuracy rates reported in this thesis are the mean values of leave-one-out k-fold cross-validation (data of single fold are used for validation and data of the rest folds are used for model training); the number of folds k is decided depending on the specific experimental design and the size of the employed database(s).

3.3 Conclusion

Three employed databases: two publicly available and one self-collected were described. The standard metrics used in the thesis to evaluate the proposed system were introduced as well, which related to both identification and verification scenarios.

Next chapter (Chapter 4) is devoted to some preliminary investigations of using EEG signal for biometric recognition.

Chapter 4

Factors Affecting Experimental Investigation

Using EEG signal for biometric recognition is still in its relatively early stage (the earliest record appeared in 1999 [5]). The characteristics of EEG signals dictate that there are several factors need and probably can only be investigated by exhaustive tests, separately. Certain such important factors are described in this chapter: the number of employed electrodes and the optimal positioning of the electrodes are investigated in Section 4.1; as the Wavelet Transform is used for feature extraction, the wavelet function(s) to be employed need to be optimised; along with the investigation of wavelet-based noise removal methods (devoted in Section 4.2); the segmentation and overlapping as two conventional steps are addressed in Section 4.3 and Section 4.4. The comparison of different conventional features for EEG-based biometric person recognition is presented in Section 4.5. The classifiers used for recognition/classification are investigated in Section 4.6, as well as the effectiveness of different stimuli. Other delicate matters such as the sensitive frequency bands (Section 4.7), optimal number of observations in the training set (related to recording length and window size during segmentation, Section 4.3 and Section 4.8) and the quality of the data provided by subjects for enrolment (Section 4.9) can be also quite influential in EEG-based biometric scenario. The following subsections detail some experimentally investigations on these important issues. The Conclusion is presented in Section 4.10.

4.1 Number and Position(s) of the Employed Electrode(s)

Most of the EEG headsets are equipped with multiple electrodes; many of these sophisticated sensors are medical-grade probes and user-unfriendly (wet sensors). In order to better facilitate the EEG signals for biometric purpose while using these headsets, reducing the number of the employed electrodes becomes a necessity. Intuitively, with the number of included electrodes reduced, degradation of the recognition rate is to be expected. Therefore, it is important to investigate the

influence of reducing the employed number of electrodes to find the best trade-off between the easy-deployment requirement and the successful recognition rate. Several electrode adoption schemes were experimentally investigated in the study, from employing multiple sensors to single sensor, along with their (its) arguably best location(s). The following subsections describe several electrode locating schemes included in the research.

4.1.1 Scheme I

The following graph (Fig. 4.1) illustrates the first electrode positioning scheme (Scheme I).

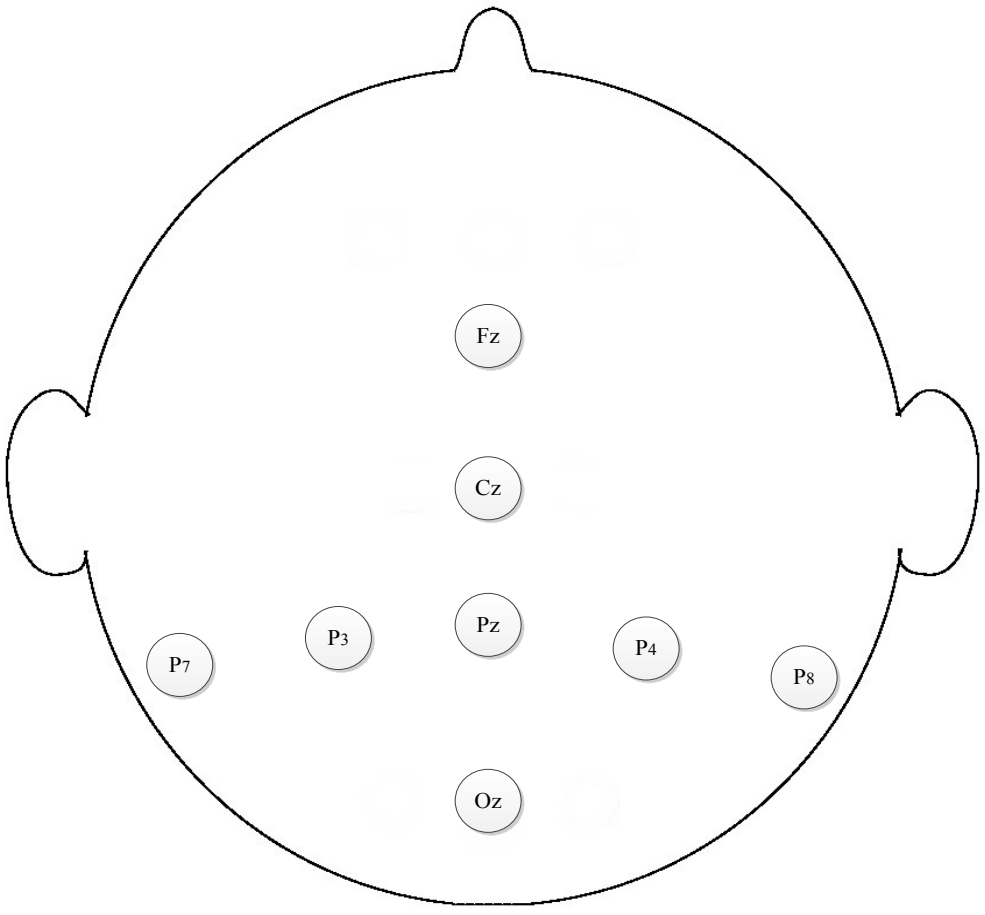


Fig. 4.1 Positioning Scheme I

The positioning Scheme I was developed to investigate the regions of interest based on the stimulus involved in the MM/I dataset. The main tasks involved in MM/I

dataset were triggered by motor movement/imagery-related stimuli. The eight selected electrode positions mostly cover both the motor cortex and visual cortex, as it is possible the brain also picture imagery images during performing the motor tasks. Data provided by these electrodes (Fz, Cz, Pz, Oz, P3, P4, P8, P7) were used to evaluate the biometric recognition performance.

The recognition performance achieved by each of these 8 electrodes was individually evaluated using a nearest neighbour classifier (1-NN). Data of 12 subjects from MM/I (S1-S12) dataset were used for the evaluation in an identification scenario. The results indicated that the electrodes located in the area of parietal lobe and occipital lobe (Pz, P7 and P8) yielded better overall performance (for different mental tasks) than most of those in the frontal lobe area (such as Fz) (Fig. 4.2 and Table 4.1).

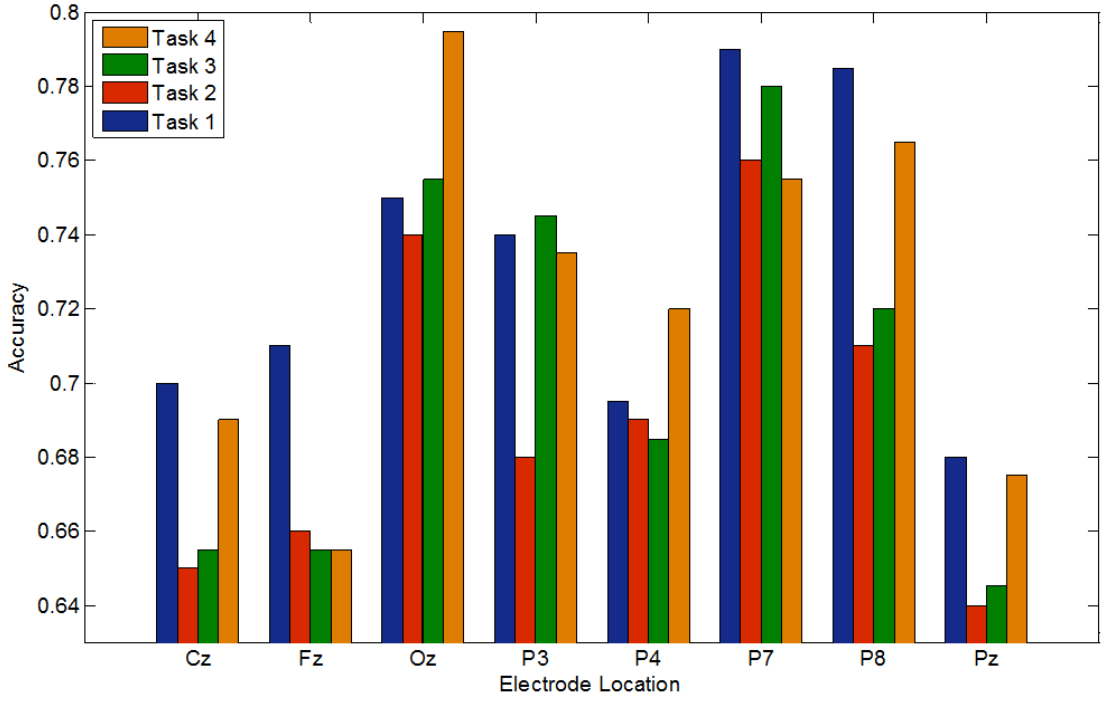


Fig. 4.2 Individual Electrode Performance in rank-1 Identification

The performance of various electrode combinations were explored as well [105]. It is noticed that the Cz and Oz electrodes generally contributed to the combinations with better performance (Table 4.1). A series of verifying tests were conducted to evaluate the performance of channel pairs: Cz-Oz, Fz-Oz, Cz-Fz, P7-P8 and P3-P4 (Table

4.1). The Cz-Oz pair produced the best average classification accuracy for a two-electrode system: 92.5%. Generally, if four best electrodes were to be used, those should be Cz, Oz, P7 and P8; if three electrodes were to be used, those should be Cz, Oz and P7 (or P8) and if only two electrodes were to be used, Cz and Oz shall be the choice. It should be mentioned here, though it is clear in Fig. 4.2 the Cz location alone does not provide the best performance, but its combination with Oz location electrode provided the highest correct recognition rate, as it is indicated in Table 4.1.

Table 4.1 Two-electrode combination identification performance comparison

Channels	Cz-Oz	Fz-Oz	Cz-Fz	P7-P8	P3-P4
Accuracy	92.5%	90.0%	86.2%	79.2%	78.8%

As the data from each single electrode contains 18000 samples which make up 20 windows before any overlapping, so one window contains 900 samples and lasts 5.625 seconds. Using two runs of data (4 minutes) for training and the other separated run for testing, the performance of each window (measurement) in the testing run was measured independently to simulate the identification and this test was conducted 20 separate times to see the variations in recognition performance. These tests were explored using the best different electrode combinations identified earlier, including the eight electrodes, best four electrodes and two electrodes mentioned above separately (Fig. 4.3).

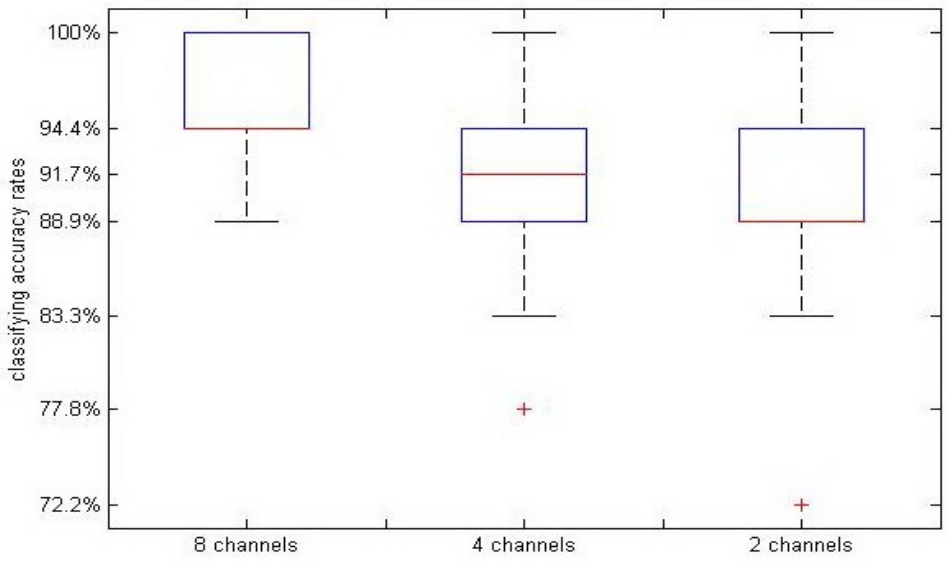


Fig. 4.3 Range of recognition rates when testing in 5.625 seconds (20 windows for each column)

From Fig. 4.3 it seems that with the decreasing number of the electrodes, the ranges of the classification accuracy rates scatter wider, the worst performance of each combination also degrades: the worst performance is about 72.2% when only 2 electrodes (Oz and Cz) are used. The results in Figure 4.3 are the performances of using the data from 18 subjects (S1-S18) and tested for 20 times: indeed, even using only 2 electrodes it could occasionally achieve 100% of classification accuracy, though only two times out of 20 tests [105].

4.1.2 Scheme II

According to the results obtained from the data captured by the electrodes in scheme I, it was found the electrodes combinations in the centre line of the scalp seems to generally provide better performance than those locate in other scalp regions. The reason behind such evidence is not quite clear, but it might have to do with the symmetric shape of the brain, which casts certain effect on the current flow waves on the surface of the scalp [108]. To further investigate the impact of the electrodes locate in the centre line of the scalp, positioning scheme II was proposed (shown in Fig. 4.4).

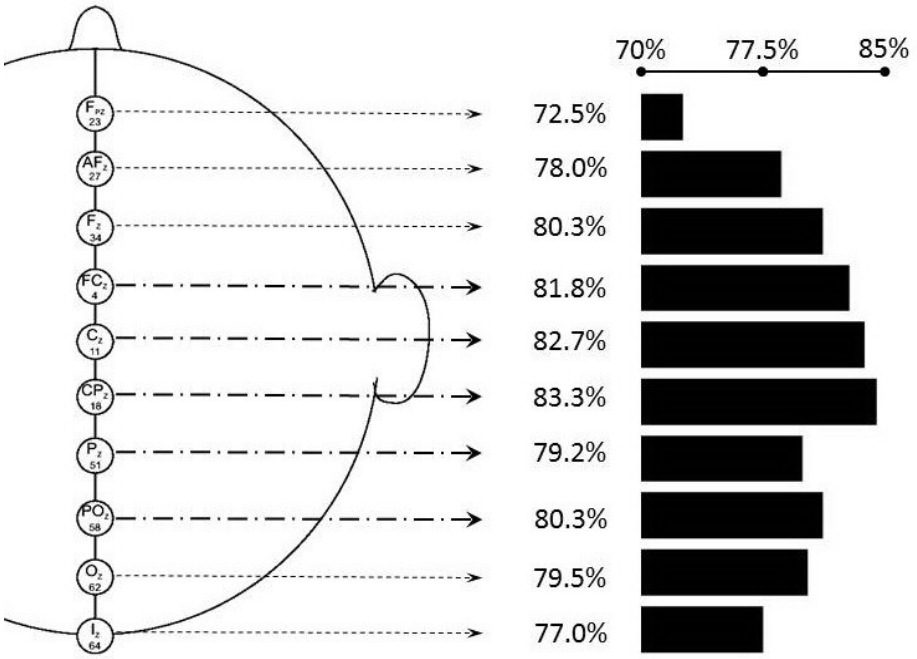


Fig. 4.4 Positioning Scheme II

The employed data volume from MM/I dataset was increased to 50 subjects and tested in an identification scenario. The percentage numbers in the Fig. 4.4 indicate the accuracies tested by the data from each selected electrode individually. Ten electrode position candidates were separately evaluated using a Linear Discriminating Analysis classifier (LDC) [76] to select a subset of these electrodes for the subsequent system evaluations. Two of the recording runs were used for training and the remaining one was used for the test. The reported results are the average of the accuracies among the three runs used in turn for testing. The dot-dash lines indicate the five selected electrodes (FCz, Cz, CPz, Pz, POz). These electrodes in combination provided the highest classification rate compared with other possible five-electrode combination schemes. However, the selected five electrodes were not individually the top-five scored electrodes amongst the ten electrodes tested: it appears that there might be spatial correlation between the signals captured by these electrodes which affect the overall performance [108].

4.1.3 Scheme III

Further investigation was done by adding the data of more subjects (the entire MM/I dataset, 109 subjects). The proposed system is trained and evaluated using data obtained from only the Task 4. The reason for adopting Task 4 (motor imagery task for both hands and feet) is that the motor imagery task might better avoid the contamination of the EEG signal by other muscle-related signals such as electromyography (EMG) signals.

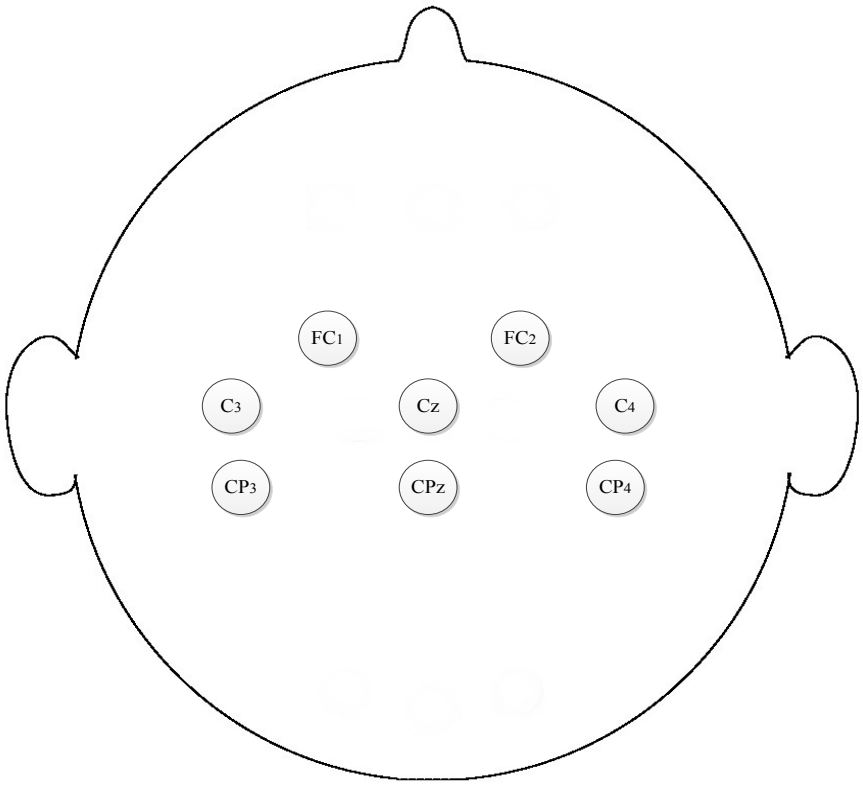


Fig. 4.5 Positioning Scheme III

Depending on the characteristic of the considered task and results of previous schemes (scheme I and scheme II), the employed electrodes from analysis was clustered in the somatosensory cortex (FC1, FC2, C3, Cz, C4, CP3, CPz, and CP4). The positioning scheme III is illustrated in Fig. 4.5. Data of the second run (R2, 120 seconds of recording) was used for testing and the rest two runs (R1 and R3) were used for training the system. The correct recognition rate reached as high as more

than 90% for data of 109 subjects by applying a novel quality filtering algorithm (to be introduced in Chapter 7 [106]).

4.1.4 Scheme IV

To explore the limitation of using EEG signals for person recognition, the involved electrode number is reduced to one. In scheme IV the performance of using the data obtained from the single electrode position (Cz) is investigated. Data of the middle run (R2) in MM/I database and the data of VEP database (only has one session) were employed. In Chapter 6 two novel features based on enhanced Empirical Mode Decomposition (EMD) were proposed in [7]. The detailed description of the new features will be presented in the feature extraction chapter (Chapter 6). Comparisons of the proposed system performance (using the EMD-based novel features) with other reports under similar scenarios are illustrated in Table 4.2.

Table 4.2 Comparison of different systems; this table also appears in Chapter 6 for feature comparison.

VEP Database	Proposed System	Yazdani *et al*. [74]	Brigham *et al*. [56]	Huang *et al*. [66]
Number of Electrodes	1	64	64	64
Number of Subjects	118	20	120	116
CRR	95.88%	100%	98.96%	95.1%

4.2 Noise Removal

In order to reveal the effectiveness of these transforms in realistic applications, the input signal with high enough Signal-to-Noise Ratios (SNR) is to be expected. In practical, EEG signals are often contaminated by other bio-signals such as eye blinks, heart pulse, and ambient electrical noise [117]. To achieve acceptable SNR, pre-processing sometimes becomes a necessary stage in EEG based pattern recognition.

The pre-processing of raw signals is divided into three steps: 1) noise removal, 2) segmentation and 3) window overlap, which may take place during segmentation. Not all of these steps are necessary and depending on the particular system or experimental design certain step(s) can be omitted. The following three sections

(Section 4.2.1-Section 4.2.3) are devoted to describe the main three pre-processing step mentioned above, supported by experimental results.

There are many noise removal methods/algorithms available in the field of DSP; one conventional method of removing the unwanted data buried in the raw signal is to perform Fourier Transform and apply band-pass filtering, hence only the data within the interested frequency range is preserved. Many classic filters, such as Butterworth and Elliptic filters have been implemented for EEG data [41]. Unfortunately, certain types of noise (eye blinks and ECG signals, for example) share the same frequency range with the typical EEG band [118][119][120][121](0 to 40 Hz): performing band-pass filtering alone in the frequency domain does not seem to be quite effective in removing certain types of bio-signals buried within EEG signals. For this reason, classic band-pass filters were usually employed by researchers to remove the Gibbs phenomenon (ripples of the output signal after Fourier Transform) around the cut-off frequency of the signal, and serves as a preliminary step of pre-processing [122].

Considering the complex content of the raw EEG signals, some researchers tried to remove the irrelevant signal components by employing blind signal separation (BSS) algorithms [123]. Independent Component Analysis (ICA) is amongst the most popular techniques for EEG data analysis. The general definition of ICA is described as follows [124]:

The data are represented by the random vector $x = (x_{1,\dots,}x_m)^T$ and the components represented as the random vector $s = (s_{1,\dots,}s_m)^T$.The task is to transform the observed data x, using a linear static transformation W as $s = Wx$, into maximally independent components s measured by some function $F(s_1, \dots, s_n)$ of independence.

Theoretically, in order for ICA algorithm to work, the mixed-up signals are presumed to be independent and stem from different sources: it is assumed the EEG signals captured by multiple electrodes are independent from each other as well. However, such an assumption may be debatable: it is believed the neurons of a normal brain are connected and related: one stimulus to one region may have certain influence on other regions of the brain (in a wave form). ICA forces the likely correlated signals to be independently separated and generates a series of newly reconstructed signals.

In general, ICA cannot identify the actual number of signal sources. Furthermore, there has been obscure in clearly justifying that the reconstructed "clean" signals are the original signals. Given these limitations, most of the researchers simply use ICA to remove eye blinks noise and Gaussian noise, which stem from different and independent signal resources [125][126]. Despite of these mentioned problems, the most challenging part for ICA to work lies in the process of identifying the resulting signal of interest. Among the generated series, normally it is up to the researcher's experience to identify the clean EEG signal series, which sometimes can be quite difficult [124].

Indeed, according to our preliminary experimental tests, using ICA for noise removal and feature extraction was unable to obtain acceptable recognition results. From the perspective of practical application, ICA is not an optimal option either: the algorithm contains a series of looping steps to compute the independent components, for a very limited amount of data (10 subjects' data each lasting for 2 minutes, in one of the preliminary investigations in this research), the processing took much longer time than expected (several minutes, Intel Core i7), for noise removal only.

In order to deal with the non-stationary nature of EEG signal while at the same time reduce the computation time during processing, some wavelet-based algorithms were investigated in this research. Unlike BSS algorithms, wavelet methods do not claim the separation of the signals from different resources, indeed the resulting signals after transform probably are still mixed up with noise. But it is expected that the wavelet approaches may preserve most of the information-bearing part of the original signals, thus improves the signal-to-noise ratio.

One distinguishing advantage of wavelet-based de-noising is its multi-resolution analysis characteristic: it is possible to closely observe a particular wavelet band in different levels (resolutions), such flexibility increases the odd of success for noise removal. In this study, three wavelet-based noise removal methods are investigated. Additionally a novel hybrid de-noising method is also developed. The performances of these methods were compared in an identification scenario using EEG data from

25 users (S1-S25) of the publicly MM/I dataset as a preliminary research to test the algorithms. The following subsections introduce these methods.

4.2.1 Wavelet Coefficients Thresholding

The wavelet shrinkage approach for noise removal, proposed by [127], assumes a noise model for its operation. Regression models are often used to recover the underlying signal which is mixed up with the noise. Such models may be expressed in the form: $y_i = f(x_i) + \varepsilon_i$, where y_i is the mixed signal, $f(x_i)$ is the "clean" signal function and the ε_i is the noise function ($i = 1, \dots, n$), ε_i is assumed to be a Gaussian white noise with unknown variance σ^2 [128].

The wavelet shrinkage approach assumes that the useful information is mostly represented by the approximation coefficients (low frequency part of DWT) generated by wavelet decomposition. The other set of coefficients that are produced after wavelet transform, namely the high frequency or detail coefficients are regarded as noise. However, if the whole of the detail coefficients were removed, there may be loss of some useful information. The "hard" threshold strategy was adopted for our experiments, which performs a "keep or kill" policy on the wavelet detail coefficients using the mini-max principle [129][130]. Two wavelet coefficients thresholding schemes were tested: global threshold and level-dependent thresholds. The global threshold scheme calculates a single threshold based on the minimum-maximum estimation, whereas level-dependent thresholding allows for specifying a separate threshold for each different decomposition level.

4.2.2 Multivariate and Multi-scale Principal Component Analysis

As the selected five electrodes (Scheme II, Fig. 4.4) are closely positioned with the same sampling frequency, some noise components of the signal may be correlated and the signal quality could be improved by removing this correlation noise [130]. However, the wavelet shrinkage approach [127] assumes only independent Gaussian noise. To remove other types of noise, one option could be performing principal component analysis (PCA) [131] on the wavelet coefficients to eliminate the cross-

correlation effects among electrodes. The multivariate de-noising method, proposed by [132], is such a method that combines wavelet decomposition and PCA: firstly the wavelet decomposition is performed; next the PCA is applied to the approximation coefficients for de-noising. After reconstruction by inverse wavelet transform, the PCA is applied again to the signal.

According to [133], this Multivariate Analysis is particularly suitable for stationary signal de-noising. Therefore, this method may not perform well for the EEG data used for the evaluation of the proposed system [134]. The experimental results presented in the following section verified this conjecture; multivariate analysis for de-noising resulted in a degradation of performance with a worse result than without any noise removal (to be illustrated in Fig. 4.6).

The multi-scale PCA de-noising method [135] applies the PCA algorithm to both the approximation and the detail coefficients, hence may better remove the correlated noise. In the results reported in Section 6.2.2 this method indeed has a good performance as indicated by the identification rate achieved.

4.2.3 Hybrid De-Nosing Method

Wavelet shrinkage de-noising is good at removing white noise that might be generated from electrical equipment and sensors during the EEG recording process. Since five closely related electrodes were used for data capture, spatially correlated noise may also affect the quality of the signal: this kind of contamination might be alleviated by applying multi-scale PCA analysis. A novel strategy is proposed which combines both wavelet coefficients thresholding and PCA methodologies to further de-noise the raw signals. The proposed method is as follows:

1) Apply Discrete Wavelet Transform (DWT) to the raw signal up to Level 5 with the sym8 (Symlets order 8 [136]) wavelet function, using minimum-maximum rule to estimate the mean square error with a “hard” level-dependent threshold, and reconstruct the signal after thresholding.

2) Apply multi-scale PCA analysis to the thresholded signals to further remove the spatially correlated noise by preserving only some of the (uncorrelated) principal component vectors.

Five wavelet-based de-noising methods were compared using data from the first 15 subjects (S1-S15) in the MM/I database. By applying the proposed hybrid method, the identification rate could be improved by more than 5.5% compared with no noise removal (Fig. 4.6). The level-dependent wavelet shrinkage provided better performance than the global threshold shrinkage method; it appears that the characteristics of noise in each decomposition level are different and better treated separately. The proposed method aims to remove both the independent Gaussian noise and the correlated noise, and indeed it appears to result in better performance than all of the other methods investigated.

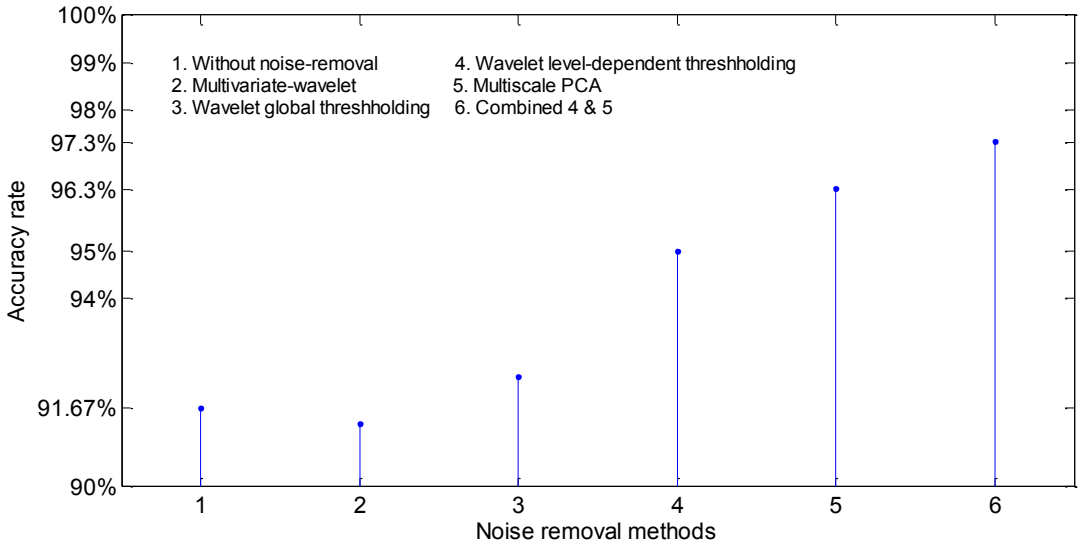


Fig. 4.6 Comparison of the de-noising methods based on biometric identification score

To further investigate the sensitivity of the proposed method, the 0~50Hz band-passed signal was divided into three frequency ranges: 0~10Hz, 10Hz~20Hz and 20Hz~50Hz. The training and testing data for 25 subjects (S1-S25) were based on their respective frequency bands. Fig. 4.7 shows that the proposed method is more effective in the higher frequency range. The performance within 0 to 10Hz even suffered some degradation. The low frequency range of the signal might not contain significant noise and may not need to be de-noised. Evidence in 0 to 10Hz also

indicates the wavelet-based de-noising method indeed could not extract the pure EEG signals out of the raw signals, better overall performance can be expected if the frequency range (0 to 50Hz) were divided into more bins based on the typical bands (delta, theta and alpha etc.) of interest for EEG signals, as oppose to the current experiment.

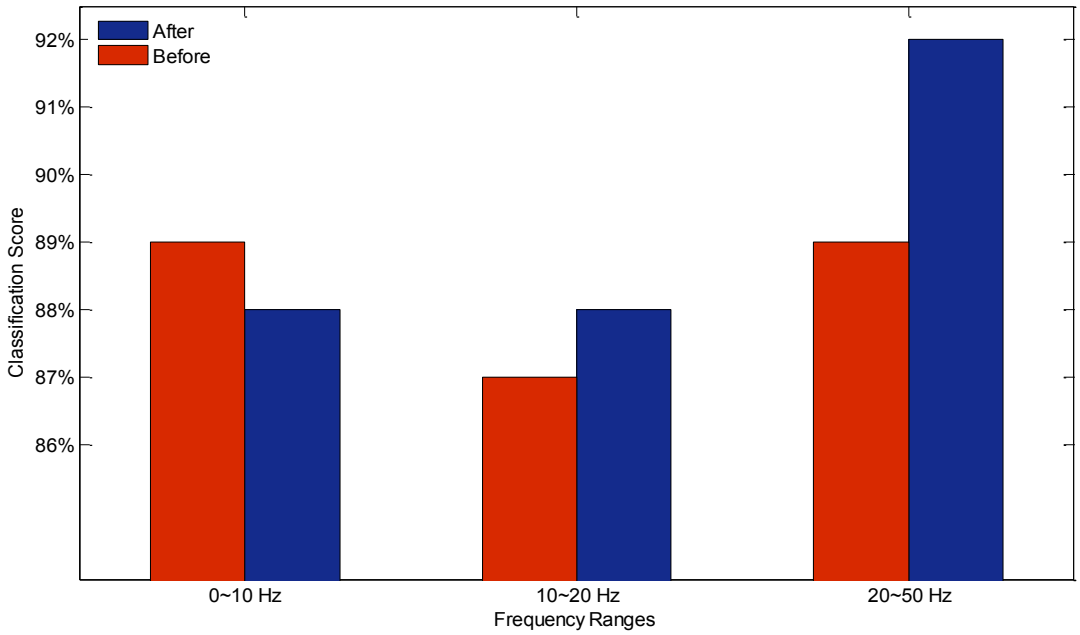


Fig. 4.7 The sensitivity of the proposed noise-removing method in different frequencies

4.3 Segmentation

After the raw signal noise removal, there is usually a segmentation step. A continuous recording potentially contains multiple events of interest which are distributed within the signal. Segmentation serves at least two purposes: 1) to break the long series into pieces hence making it possible to isolate the events of interest; 2) normalize the number of the samples for feature extraction.

In this work, a series of investigation on the optimization of segmentation based on Wavelet Transform were conducted. The raw signal was segmented into several non-overlapping windows after de-noising. The MM/I dataset was employed for this investigation. Originally every two minutes' recording (Run) was divided into 20

windows (960 samples per window before overlapped). After several tests (Fig. 4.8), windows with size of 4800 samples (30 seconds of recording with sampling frequency of 160 Hz) were chosen for further feature extraction. The biometric performance was found to be quite sensitive to the window size.

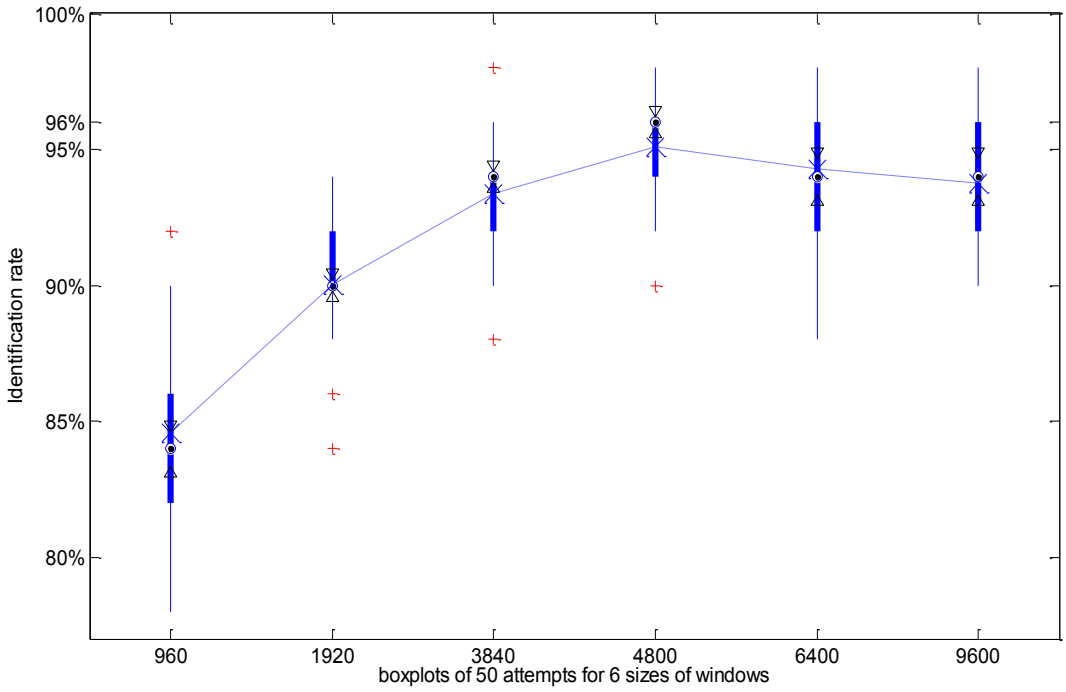


Fig. 4.8 50-fold-cross-validation tests of 6 seconds recording for windowing schemes

One of the two-minute recording runs was used to provide data for testing. Fig. 4.8 shows the accuracy rates for different window sizes. By randomly picking several 6 seconds long samples from the 2 minutes' recording, tests were performed at different pre-processing window sizes. The "×" symbol of each boxplot represents the mean accuracy rate of 50 testing attempts. With the window size increasing, the computation time decreases and the identification rate increases until it reaches its peak with a window size of 4800 samples, then the performance begins to drop to a plateau as the size is further increased. Figure 4.8 also indicates the varying consistency of results achieved with different window sizes. For MM/I database "4800 samples per window" resulting in the most stable performance as indicated by the most compact boxplot which represents the smallest variance of the 50 attempts.

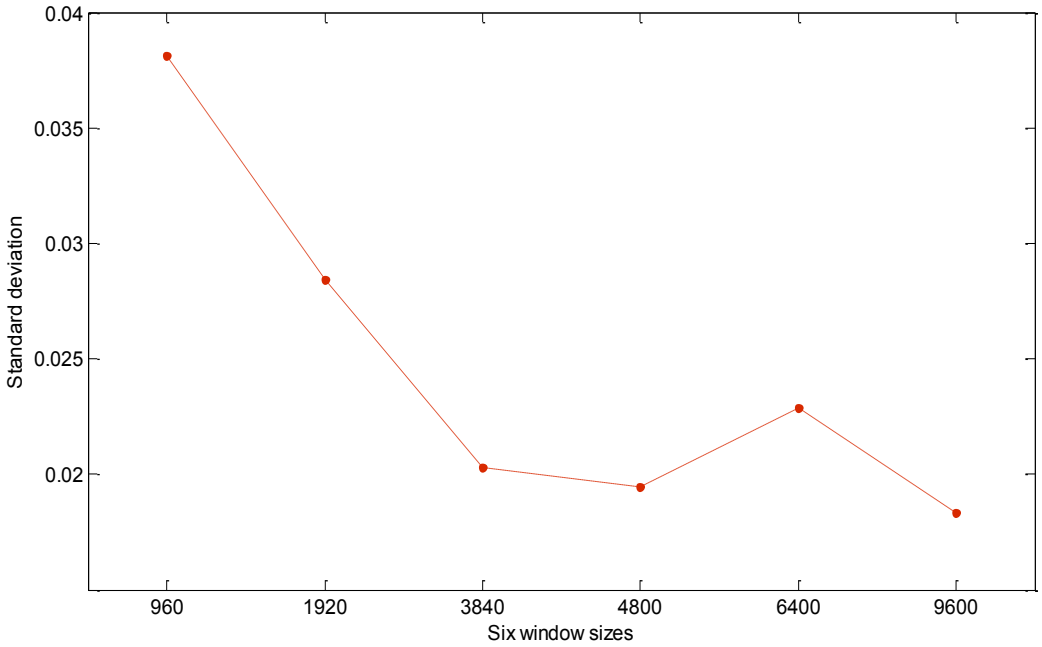


Fig. 4.9 Performance Stability of 6 Window Sizes, 50-fold-cross-validation tests

The reason for the particular trend in Fig. 4.8 may has to do with the entropy of the data after feature extraction: as a stochastic time variant series, the bigger the window length used for feature extraction the long term characteristic of the signal is more likely revealed, therefore more likely the representative trend of the signal be captured. It also can be justified by the results presented in Fig. 4.9: the standard deviations of multiple testing results with different window sizes indicate the rising trend of the performance robustness with the window size increasing. Smaller windows capture only the details of short term events, some of which may be only noisy spikes and vary a lot; whereas by employing relatively large window sizes for feature extraction, those noisy spikes are more likely be overwhelmed by longer term activity (however, this also depends on what type of features are used and what kind of mental activity is processing). It is rather difficult to find an optimal window size for all of the experimental cases (if not impossible).

The optimum window size of the segmentation is governed by multiple factors: the sampling frequency of the system is one important factor; the type of feature vector developed for classification is another. It should be mentioned that the employed

MM/I dataset was under the sampling frequency of 160 Hz, which means the optimal 4800 samples per window's segmentation scheme requires at least 30 seconds' data recording from subject in order for feature extraction. Such restriction/trade off can be quite influential in the real-life scenario. It is interesting to investigate whether there is a way to reduce the recording length for each window and keep or further improve the performance in the meantime. A novel algorithm was developed and will be introduced in the feature classification chapter (Chapter 8).

4.4 Overlapping

Overlapping is a common procedure often employed in some signal processing and pattern recognition tasks, where segmentation windows were overlapped with each other. This may provide a better chance to capture the intact events, and avoid the potential significant events be chopped up by segmentation. The optimum amount of overlap is often a matter of empirical evaluation: commonly a 50% of the window's overlap may be used [47]. With a 50% of overlap between adjacent windows, the overall number of windows is doubled, which indicates the amount of features also doubled. Therefore, the information revealed from the data is potentially increased, though the computation burden is also increased. Wavelet Function Optimization

One advantage of the Wavelet Transform (WT) over Fourier Transform (FT) is its flexibility on choosing the mapping functions. As opposed to FT which uses only sinusoidal functions in transform, there are many different types of functions available in WT. However, like other factors considered in this chapter, there is no concrete guidance on how to select the optimal function for a particular application. The only strategy for its selection seems to be exhaustively testing the available functions in different parameters.

The biometric performance of the features that derived using Wavelet Packet Decomposition (WPD) algorithm [59] is affected by the types and orders of wavelet functions used in the feature extraction stage. Three types of wavelets were examined in this preliminary investigation: Daubechies, Symlets and Coiflets and some results are shown, the effectiveness of a noise-removal method is also addressed in Table 4.3.

Table 4.3 Wavelet Functions investigation by Identification Scores

Wavelet Function	Before De-noising	After De-noising	Accuracy Increased
Daubechies 2	80.5%	87.0%	6.5%
Daubechies 4	**84.6%**	**95.5%**	**10.9%**
Daubechies 5	83.9%	93.0%	9.1%
Symlets 1	76.7%	87.0%	10.3%
Symlets 3	**84.7%**	**94.0%**	**9.3%**
Symlets 4	83.0%	90.5%	7.5%
Coiflets 2	83.5%	91.0%	7.5%
Coiflets 4	**82.0%**	**93.5%**	**11.5%**
Coiflets 5	82.3%	90.0%	7.7%

These recognition results were obtained from the data of 50 subjects from MM/I dataset following Scheme II, as the main purpose of the investigation was to compare and find the best wavelet function for preliminary feature extraction, the overall CRRs were not fully optimised. In the case of the Daubechies and Symlets wavelet families, the orders from 1 to 20 were exhaustively evaluated and the Coiflets were tested from order 1 to 5 [137]. The results indicated that as the wavelet order increase the impact on performance became less significant. The optimal wavelet function for WPD feature extraction was chosen as the Daubechies 4. It should be clarified that though the results in Table 4.3 also included the effectiveness of a novel noise removal method, the accuracies' trend and the optimal wavelet function have been clearly shown. Daubechies 4 wavelet is selected for the MM/I dataset and will be used for all the wavelet-related methods throughout the thesis.

4.5 Investigation of Some Conventional Wavelet-based Features

Feature (vector) is the key factor in pattern recognition and its many successful applications. Conventional wavelet-based feature extraction in EEG biometrics usually is based on utilizing the resulting coefficients after either the Discrete Wavelet Transform (DWT) or the Wavelet Packet Decomposing (WPD) [137]. Wavelet Transform (WT) can generate large number of coefficients due to its multi-scale decomposition functionality. As informative as these coefficients maybe, it is not ideal to utilize all of them as features due to the "curse of dimensionality" [138] (feature vectors with too high dimensionality to be classified correctly).

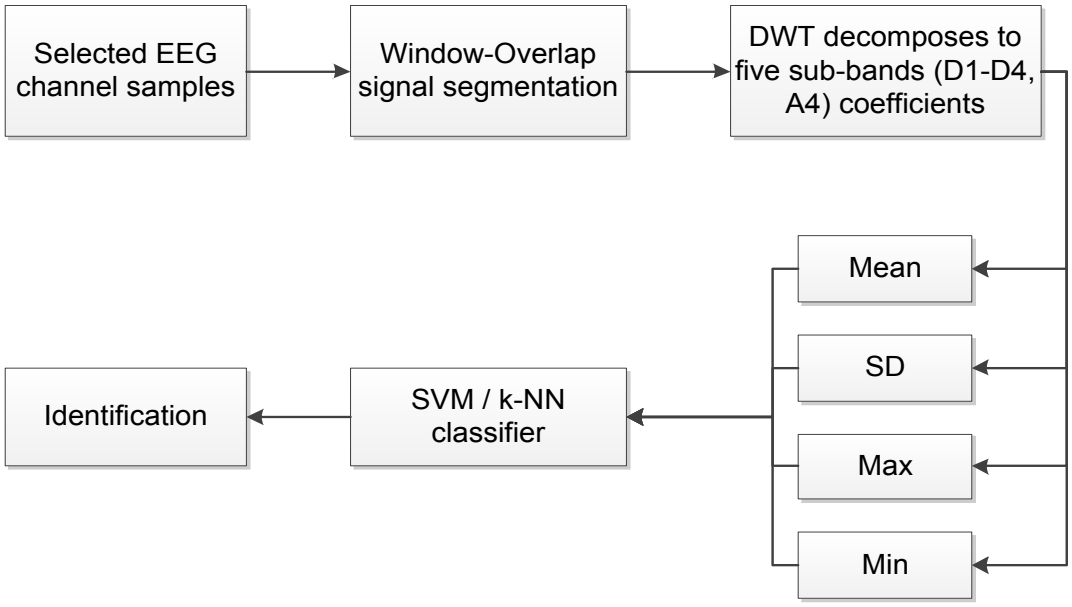


Fig. 4.10 Tested features in the experimental system

Several simple mathematical tricks could be employed to reduce the dimensionality of the resulting feature vector and (arguably) preserve most of the useful information of the signals (Fig. 4.10). In this investigation, four different tricks were employed to reduce the dimension of the resulting wavelet coefficients: Maximum, Minimum, Mean and Standard Deviation (SD) of the coefficients from multiple bands. These features were computed for each and every window after WT.

For every window, after applying the Wavelet Transform, these four features were calculated for each sub-band (D1-D4, A4), thus every window for each electrode

yields 4×5=20 features. In summary, for every subject (12 subjects based on scheme I), 20×8=160 features (eight electrodes, each electrode generating 20 measurements) are calculated from each window. The four different features were analysed separately in order to make a comparison between them. The investigation of the features follows the Scheme I electrode positioning. The results obtained according to the above-stated strategy suggested the SD of the coefficients is the optimal feature amongst the four feature candidates.

4.6 Classifiers and Tasks

Amongst all the important factors in EEG-based pattern recognition, the choice of the employed classifier is one of the most influential and unpredictable factors: Provided the same available feature(s), different classifier can lead significantly different recognition performance, it depends on the distribution of the available data in the feature space. Here is the difficulty: there is no a priori knowledge can be obtained to predict the distribution of the feature points before the decision making, except for the boldly assumption of all the features follows an existing mathematical model (normally distributed, for example). Thus it seems there is no guidance on how to select the classifier for various scenarios. On the other hand, instead of depending on the predicting rule (if there is any) to select the optimal classifier, it is often much more convenient to simply test throughout all the available classifier candidates in simulation and deploy the best-performed classifier in its real-life scenario.

The selection of the classifier in this thesis also follows this simple and dependable strategy: multiple classic classifiers were tested by the databases and the optimal ones were selected based on their respective experiment results.

One of those experiments was devoted to compare the performance of *k*-Nearest Neighbour classifier (*k*-NN) and Support Vector Machine (SVM) using MM/I dataset. 12 subjects (eight electrodes positioned as Fig. 4.1) are used for evaluations in an identification scenario. In this specific experiment, multiple feature vectors of 40 measurements (each measurement corresponds to 5 seconds) with 80 dimensions (mean and SD of wavelet coefficients per window as feature) are generated for one

single task (12 classes). Through an empirical evaluation, the proximity parameter P for the Euclidean kernel (for SVM) was set to 2 in order to yield the best performance [105].

During the testing stage, for each task all the analysed windows from one randomly selected run were treated as test data; The best average classification rate produced by SVM was 97.4% for Task 4 (imagery task), and the worst one was 75.7% for Task 3 (movement task); while for the k-NN classifier (with k=1) the best classification rate is 85.09% and the worst rate is 74.12%, belongs to Task 4 and Task 3 respectively (Fig. 4.11).

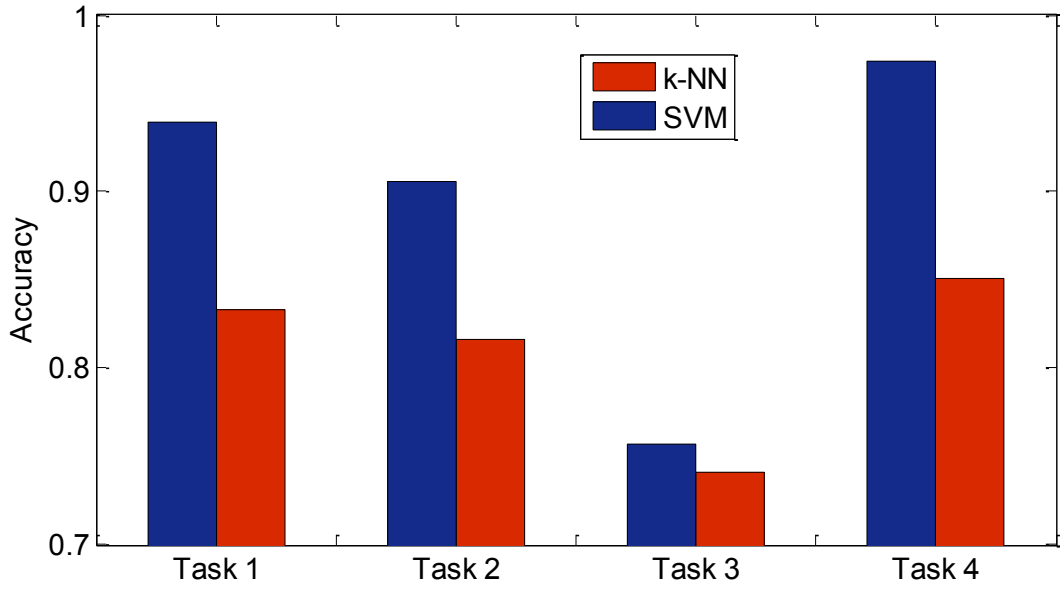


Fig. 4.11 Comparison of the recognition performance for the four stimulus tasks

The explored four stimulation tasks could be divided into two groups: motor movement and motor imagery, it seems that the imagery tasks generally outperform the movement tasks when it is used to provide EEG signals that generate identity information (identification scenario). Also based on the results from Table 4.4, for different tasks the optimal number of neighbours for the k-NN classifiers are not all the same: They are 7-NN, 1-NN, 3-NN and 1-NN from Task 1 to Task 4 respectively. This suggests that the imagery tasks (Task 2 and Task 4) may be producing more compact feature sets than movement tasks (Task 1 and Task 3).

Table 4.4 Performance compared between SVM and *k*-NN (12 subjects)

Tasks	1	2	3	4
SVM	93.9%	90.6%	75.7%	97.4%
k-NN	7-NN	1-NN	3-NN	1-NN
Highest Rate	83.3%	81.6%	74.1%	85.1%

The optimal combinations of electrodes were selected for the current database and selected individuals (12 subjects). By further analysing the performances achieved using two electrodes it appears that if only one electrode was to be retained, the best choice may be the Oz or the Cz electrode. Exploring and optimizing the performance of the single EEG electrode for biometric recognition will be described in in feature extraction chapter (Chapter 6) with more details.

4.7 Frequency Bands

Several EEG frequency bands were identified and recognised by researchers. There are five typical bands cover the frequency range from <4Hz to >32Hz: namely delta (<4Hz), theta (4Hz-7Hz), alpha (8Hz-15Hz), beta (16Hz-31Hz) and gamma (>32Hz) [139]. These bands are often related to certain sensory/mental activities/statuses, however, those statuses/phenomenon are generally based on experimental evidence and analysis without concrete justification. Moreover, the effectiveness of EEG's frequency bands on biometrics is still obscure and controversial.

A series of investigation were conducted for the interested frequency bands range from >0 Hz to 50Hz, which cover all the conventional EEG bands. The MM/I dataset was employed. The multi-scale analysis characteristic of WT was utilized and the EEG signals of 109 subjects were decompose till level 5 to specify the bands of interests. The wavelet decomposition tree illustrated in Figure 4.12 was based on Daubechies 4 wavelet function, the bold bands were preserved and the coefficients of which were used to compute the SD feature.

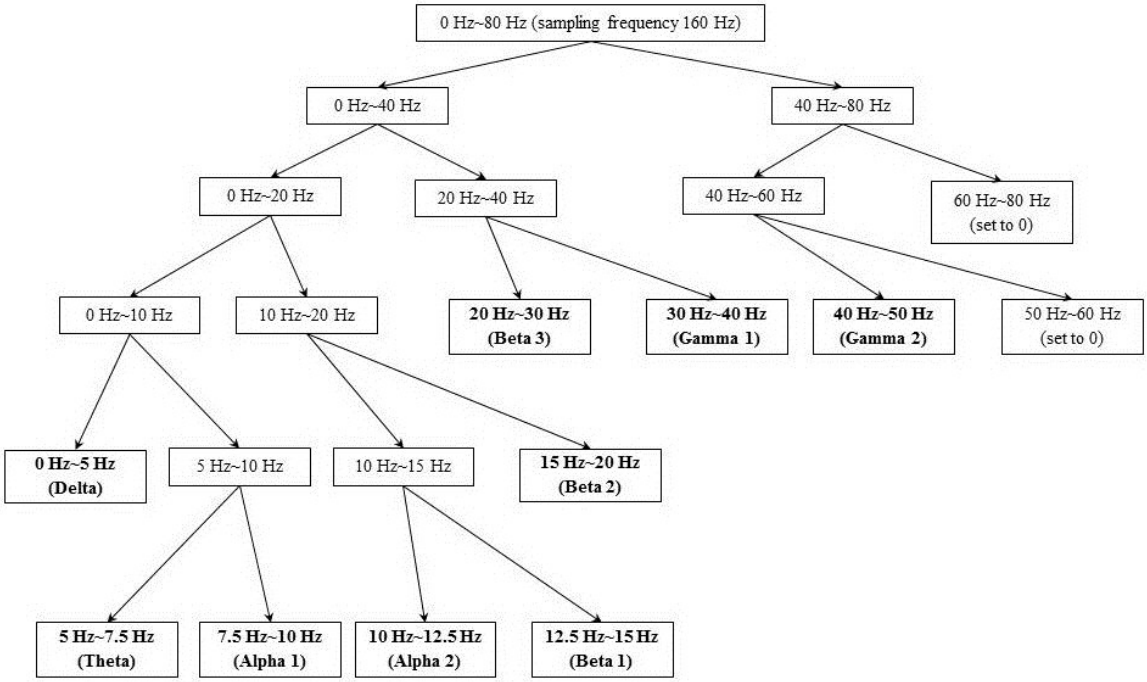


Fig. 4.12 Wavelet packet decomposition for frequency sensitivity analysis

Different combinations of frequency bands were investigated to establish the respective contribution of each band to identification performance. Nine bands were separated into three groups with each contained three bands. The performance of individual bands and groups are depicted in Fig. 4.13. The electrodes positioning followed the Scheme III (eight electrodes).

It appears that the middle frequency range (5 Hz to 20 Hz) contributes the least biometric information (2^{th} to 6^{th} columns), whereas the 0-5 Hz band along with the 20 Hz-40 Hz bandwidth provides much better performance. These frequency bandwidths were then further combined as depicted in right part of the Fig. 4.13. While the frequency range of 0-10 Hz combined with the 20 Hz-50 Hz range (14^{th} column) reached the performance level almost as good as that achieved by the combination of all the bands (16^{th} column).

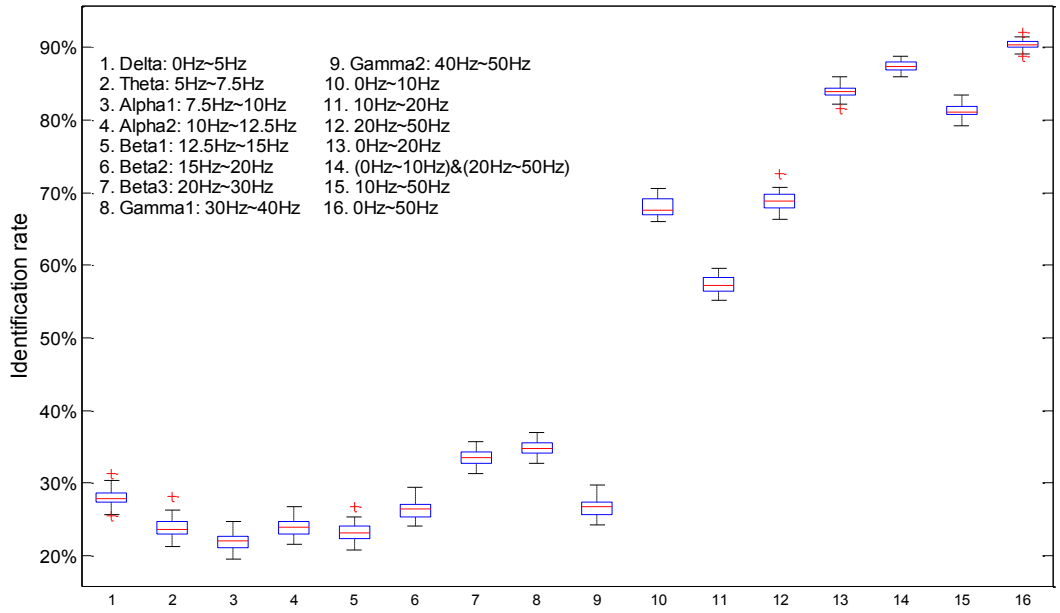


Fig. 4.13 Performance for 50 attempts, Run 1 & Run3 for training, Run 2 for testing (scheme III)

4.8 Impact of Training Data

As a biometric modality, EEG signals may be debatable in whether it should be considered as physical trait or behavioural trait; based the research results in this thesis, it is more likely belongs to the behavioural modality category. One of major reasons is its rather unstable intra-class biometric performance: the impact of separated recordings (even with short time interval) is significant and could not be ignored. Take the MM/I dataset for example, though it contains three separated recordings for each task per subject, each of these "run" in fact could not be considered as a session due to the very limited time interval between recordings (a couple of minutes), and most likely without the removal of the headset. However, even with such limitation, still there is considerable impact can be observed according to the investigation.

The identification performance using MM/I dataset was compared while data of different runs were used for training the classifier: Table 4.5 shows the results from three different schemes for partitioning training and test data. Combining Run 1 and Run 3 for training provided the best identification performance. The reason behind such variance might lie in the different quality of each recording Run; the R1 & R3

for training with R2 for testing appears to be a better model for the classification, which has roughly 5% ~ 7% better average recognition accuracy for 109 subjects.

Table 4.5 Performances of different training set combinations

Runs for Training	R1 & R2	R1 & R3	R2 & R3
Maximal Accuracy	93.55%	99.28%	94.98%
Mean Accuracy	90.20%	97.44%	92.22%

The amount of data used for training was varied to assess how a reduced training data volume may affect the recognition accuracy of the system. Define the number of observations (measurement) after segmentation of the raw signals as *N*, which was reduced in a number of steps from 80 to 1 and the identification test results are reported in Fig. 4.14.

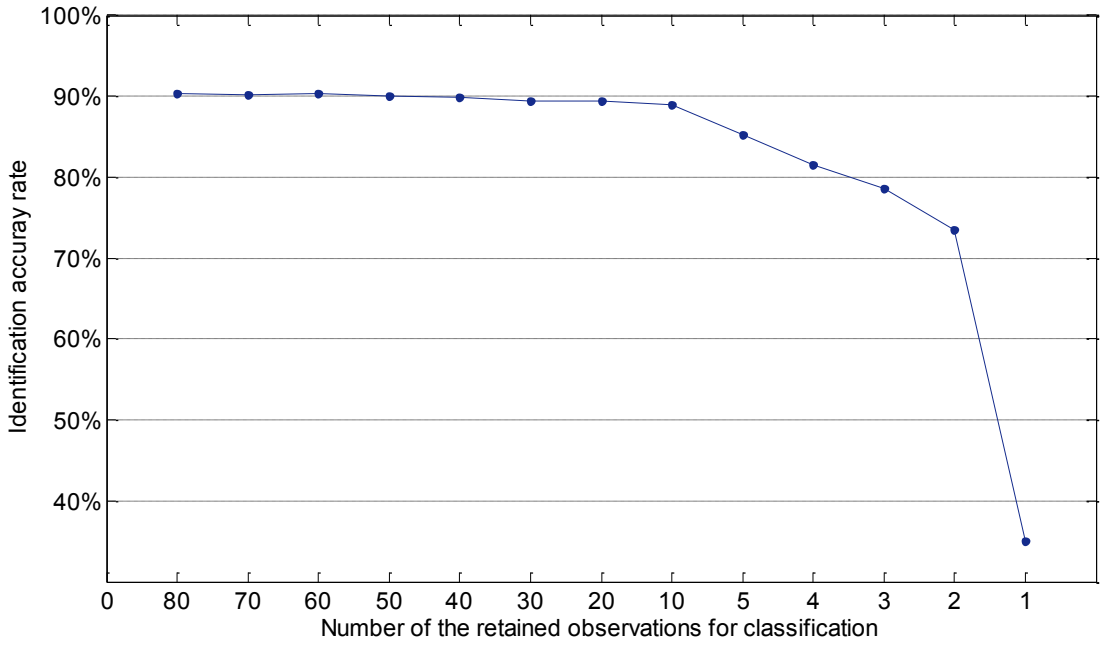


Fig. 4.14 Identification accuracy as a function of the number of training observations used for training (Scheme III)

The electrode positioning followed the Scheme III (eight electrodes). The results were relatively stable up to *N*=10. Still more than 70% accuracy rate can be achieved

while only 2 observations were kept. However, the performance significantly degraded when only one observation (6 seconds of EEG recording) was preserved for training. A compromise setting may be N=10 (10 observations) which roughly using 1.3% of the data is utilized for model training, and still achieving more than 90% identification rate for 109 subjects.

It should be mentioned that the results in Fig. 4.14 were obtained by a novel data filtering method using Sample Entropy (SampEn) to assess and preserve the most informative segments (windows). Without applying that filtering method, the performance's degradation would be much faster. The detail of the SampEn-based filtering algorithm will be described in the feature filtering chapter (Chapter 8).

4.9 Enrolment Control

In a real-life biometric scenario, it is very important to keep the quality of the database above certain level by evaluating each new user's data during the enrolment stage. The considerable intra-class variation of EEG data has been mentioned in Section 4.8 (Table 4.5); it would be good to control each new data enrolment and guarantee the new user to be added into the database is not the "black sheep" and spoil the entire training model.

An enrolment control scheme is developed for the new user's enrolment to maintain the overall quality of the database and prevent the excessive degradation of the data quality during the enrolment. Indeed, it is reasonable that with the involved number of subjects increasing, the performance of the database would degrade monotonically, provided each subject could offer the data with approximately the same quality. However, since EEG signal is non-stationary and easily interrupted, quality of the signal may vary a lot for different subject.

The proposed scheme is designed to evaluate subjects' data quality by a pre-classification during the enrolment: by utilizing an efficient classifier (LDA was chosen for this experiment), the system will do a series of classification with the number of subjects increase one by one. Fig. 4.15 depicts the overall performances before and after the subject evaluation module.

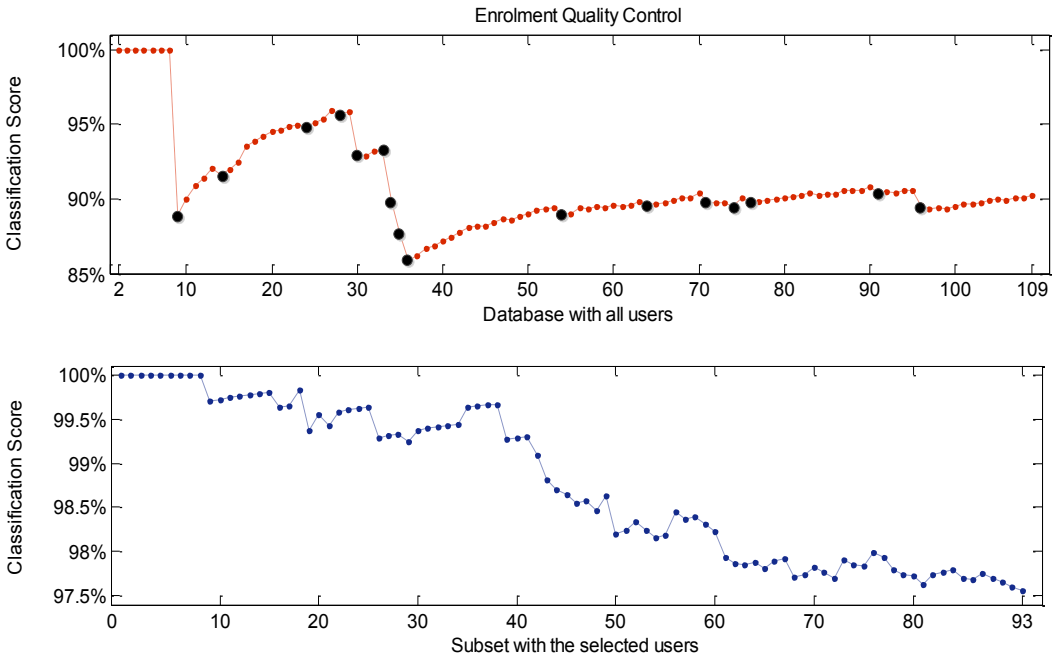


Fig. 4.15 Subject's data Quality Control

The upper graph in Fig. 4.15 shows the trend of the performances from the adopted database (MM/I, all 109 subjects), it can be seen the black points (each point represent one subject) led to performance degradations while their data added to the database for classifier training, the degradation can be found after the detected 16 subjects data joined in the database.

To automatically detect the subjects with low quality data, the performance difference between two subjects was calculated for each new enrolment, i indicate the added subject:

$$Trend\ (i)\ =\ accuracy\ (i)\ -\ accuracy\ (i-1), i = 1{:}109 \qquad (4.1)$$

The scheme (4.1) detects the value of the *Trend* and removes those that fall out of the threshold of [− 0.001, +∞) for the employed database, the lower graph of Fig. 4.11 shows the optimised performance and the gradual performance degrading process, for the acceptable 93 subjects (removed 16 black dotted subjects) the classification rate reached up to more than 97.5%. Whereas without quality control, for first 93 subjects (S1~S93) the accuracy rate is about 90% (the upper graph of Fig. 4.15). It appears that a small amount of low quality enrolments could affect the entire dataset's

modelling badly. With this simple and effective scheme, every new subject who are trying to enrol to the database will be evaluated and those who provide low quality data will be suggested to re-do the data capture.

The results in Fig. 4.15 is an illustration of the results from the proposed evaluation scheme, the threshold can be tuned depend on the specific scenario requirement and the resulting plot could be further monotonically smoothed. The blank dots indicate the detected "black sheep" during the enrolment. The proposed scheme is only suitable for the enrolling process, when the true identity is available. These results were generated by utilizing part of the enrolment data (2/3 of the available data) for model training and the rest for generating the quality scores/accuracies.

Another issue that needs to be clarified is the efficiency of the evaluation: since the testing process is repeated multiple times with the increasing of the users, the training time becomes a factor to be reckoned with and the choosing of LDA in this study is result of such consideration. Linear Discriminate Analysis is a both efficient good performed classifier for EEG signals: the time consumed for 109 subjects' evaluation was no more than one minute.

4.10 Conclusion

In this chapter a series of important factors related to EEG-based biometric system were investigated. The results/experience gained from these investigations is quite helpful in paving the way to develop the full-fledged recognition system.

The developing of EEG-based biometric system are detailed in the following a few chapters (Chapter 5 to Chapter 8), each chapter is devoted to further investigate one critical stage. The following chapter discuss the sensitivity of tasks which used to trigger EEG signals for biometric recognition.

Chapter 5

Biometrics Performance Sensitivity to Cognitive Tasks

Despite some considerable successes in using EEG signals for biometric recognition, important challenges still hinder their widespread adoption and acceptance [140], because of this the search for new biometric modalities continues. Bio-signals from brain are difficult to copy accurately (spoof) due to their non-stationary characteristics [141], which makes them appealing candidates for biometric applications. The EEG signal is a good choice as a biometric modality due to its increasing ease of acquisition through low cost sensors.

5.1 Introduction

Poulos *et al*. first proposed employing EEG signals for person identification [5][6][142]. Since then, this modality has received increasing attention for biometrics applications. Intuitively, EEG signals may be expected to contain some information unique to individuals. However, it is not clear what deliberate or involuntary mental activity would generate the best and most biometrically informative signals. This question is closely related to which scalp region should provide the signals for biometric recognition.

The mental activity or task used in the research literature to trigger EEG signals for biometrics could generally be grouped into three main categories:

1) Resting state, with no intentional mental or physical activity while with eyes either open or closed
2) Event related potential (ERP) signals, especially the P300 evoked potential [35] triggered by visual stimuli
3) Intentional mental activity(s), either through motor movement or motor imagery type

Some important research results related to these three categories have been reviewed in the Literature Review (Chapter 2) subsections.

This chapter is devoted to an investigation of the performance sensitivity of EEG-based biometric systems to the choice of mental activities used for their training and test. The work is structured as follows: Section 5.2 provides a justification for using wavelet-based features and provides details of the particular wavelet features used in this work. Section 5.3 presents some research questions to be investigated as well as the database used for investigating these questions. Section 5.4 contains the proposed experimental protocols, which are especially designed to investigate the questions proposed in Section 5.3. The experimental evaluations and the analysis of results are included in Section 5.5. A discussion of the results is provided in Section 5.6 and conclusions and suggestions for further work are included in Section 5.7.

5.2 Research Questions

The impact of using different mental tasks for generating biometric EEG signals has not been investigated. Whether or not, the type of the task affects the performance that can be achieved in biometric recognition may be an important factor and is yet to be investigated. Four specific questions are addressed in this work:

1) Does the optimal placement of electrodes vary with the movement/imagery task required of the subjects?
2) Does the type of movement/imagery task performed by subjects affect the biometric recognition performance?
3) Would training with data from one task and testing with data from another task significantly affect performance?
4) Whether combining data from different types of mental tasks for training of the system affects performance?

5.3 Motivation for using Wavelet-based Features

Similar to certain modalities in the field of signal processing (speech recognition, for example [143]), the EEG signal is also considered non-stationary [141][144]. Fourier Transform is a conventional approach in signal processing and is widely used for EEG-based signal analysis. However, its use is based on the assumption that the data

to be analysed is strictly stationary. Short-time Fourier Transform (STFT) may moderately relax this restrictive criterion: segmenting the non-stationary signal into a series of overlapped short-time frames, by assuming the data within each frame is stationary, the Fourier Transforms is applied to each of these frames separately. This approach, however, may not be able to fully capture the non-stationary dynamics of the signals' content.

The window type and length employed for STFT are important factors in successfully capturing the information content of the signal. As the length of the window used in STFT affects both time and the frequency analysis, there is a trade-off in choosing the window size: a longer window will provide better frequency information, whereas a smaller window will capture more time-variant events of the signal. However, it may be difficult to select a universally optimal window length/type which applies to every user in an EEG-based biometric scenario.

One advantage of the Discrete Wavelet Transform (DWT) is the flexibility of choosing the wavelet functions. Discrete Fourier Transform (DFT) could be viewed as a special case of DWT: rather than representing the signals by a series of sinusoidal functions, DWT decomposes the signal using a series of scaled and shifted wavelet functions, different wavelets may be used based on particular applications [145][59]. From (5.1) the flexibility of the stretched and shifted wavelets allows a better simulation of the original signals, which facilitates multi-resolution analysis.

In this work the Wavelet Packet Decomposition (WPD) transform is employed, which includes a full decomposition of the signals into multiple levels using both wavelet and scaling functions [145]. In order to maximize the use of both time and frequency properties of the signal, not only the coefficients from the lowest level but also the higher level coefficients were employed as the primary features. The EEG signals were decomposed up to Level 3 (Fig. 5.1). This allows the signal to be divided into eight non-overlapped wavelet bands.

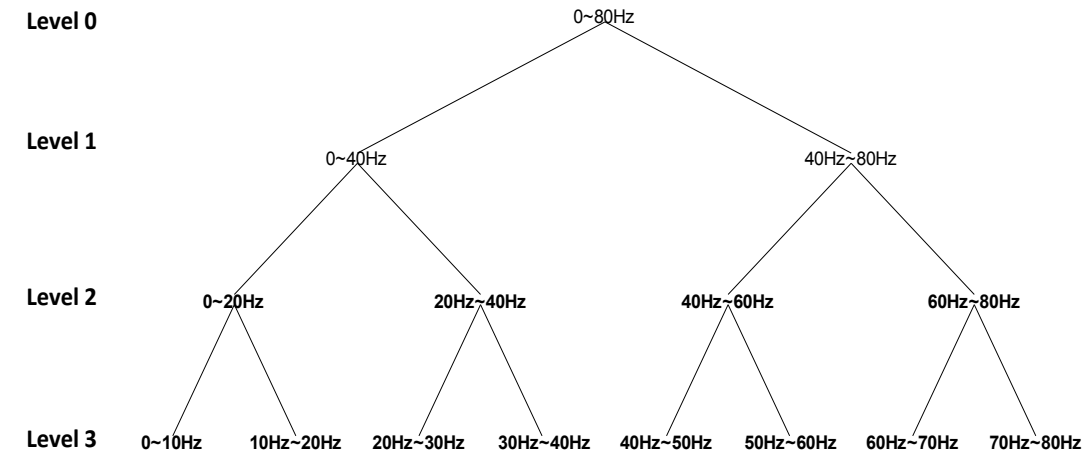


Fig. 5.1 Wavelet Packet Decomposition for the Proposed System, the bold bands are the frequencies utilized in the system

Different decomposition levels result in a series of coefficients with different lengths: the higher the decomposition level, the more frequency details are reflect in the coefficients, hence less time domain information may be retained. Therefore, the coefficients of the four sub-bands from Level 2 were retained for feature extraction as well as those of Level 3, as they may retain useful time domain properties of the signal. The Daubechies 4 was used as the wavelet function based on preliminary investigations reported in and the segmentation window size was chosen to be 4800 samples [108]. The nodes marked in bold letters in Fig. 5.1 indicate the selected bands and levels used to construct the feature vector.

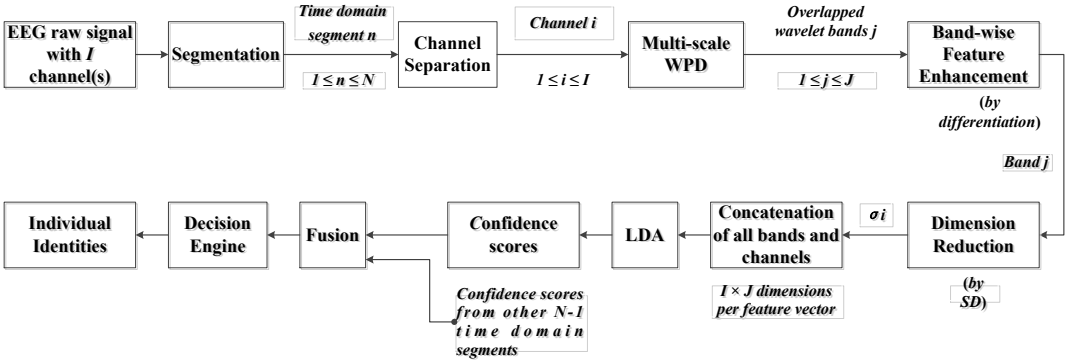


Fig. 5.2 System Diagram

As shown in Fig. 5.2 the operation is to compute the differentiation of resulting wavelet coefficients in each of the selected bands using Equation (5.1). This

operation is aimed at enhancing the temporal variations in the wavelet signal and was found to significantly improve performance, which is similar to the notion of image gradient which represents the change in intensity or colour of an image [146]. Function in (5.1) defines the differentiated wavelet coefficients of each window after the WPD.

$$F\{x\} = \frac{d}{d(x)}\left(WT_{\psi}\{x\}(a,b)\right) \tag{5.1}$$

After calculating the derivative of the wavelet coefficients, the standard deviation (SD) of each wavelet band for each window is computed. The main motivation for this step is to reduce the dimensionality of the training data: each window with 4800 samples is represented by twelve SD values representing each of the wavelet bands.

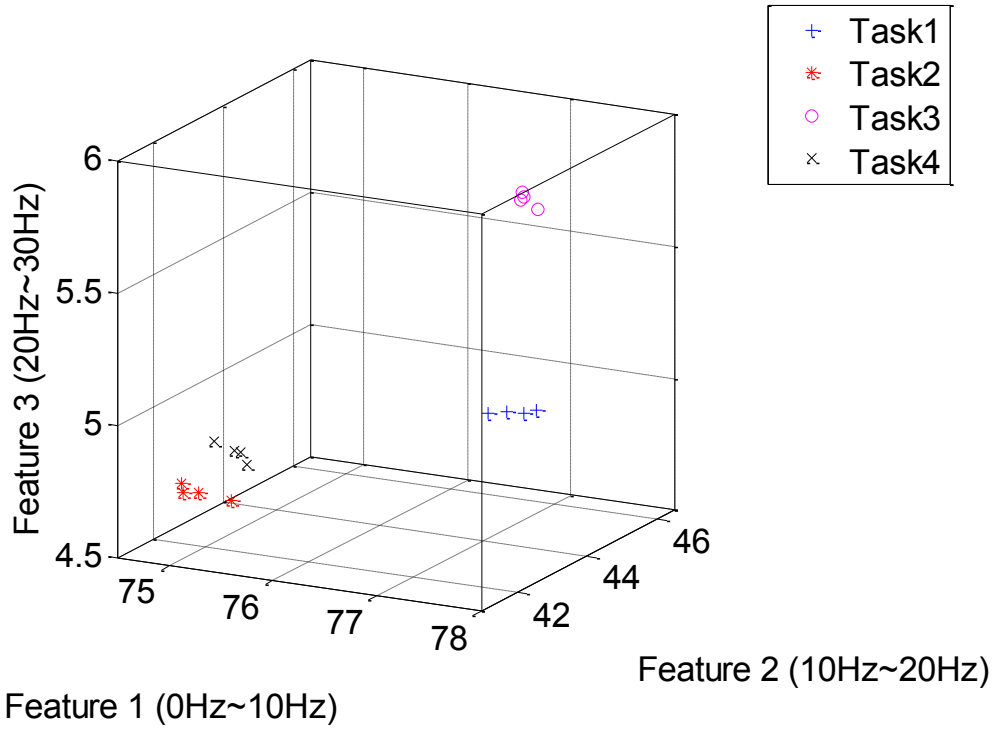
These SD vectors were used as features for training a linear discriminant analysis (LDA) classifier. The results of classification from multiple windows were combined using majority voting decision fusion. System performance for identification and verification scenarios was investigated. For the identification scenario, standard LDA was employed to train the system; for the verification scenario, however, Fisher's LDA was found to provide considerably better performance than the standard LDA. The difference between these two types of classifiers is that the Fisher's LDA does not make assumptions such as normally distributed classes or equal class covariance [77], which may explain its better performance in the one-to-one verification tests.

5.4 Experimental Protocols

Data from the "EEG Motor Movement/Imagery Dataset" (MM/I) was used for these investigations [18][110]. This dataset contains data collected using BCI 2000 (sampling frequency 160 Hz) from 109 subjects; in order to guarantee equal and sufficient recording length, 105 out of 109 subjects' data were selected for the experiments – excluding 4 subjects with shorter data recordings. Details of the task instruction can be found in Section 3.1.1 and [15].

Before conducting experiments using this database to explore the sensitivity of the biometric system to task type, it is helpful to verify that there are indeed some significant differences in the four mental/imagery tasks that it includes.

The mean of wavelet coefficients is used as a feature for task discrimination. Data of multiple subjects (first 15 subjects of MM/I dataset) were analysed and the values for the four motor/imagery tasks were plotted. As examples, Fig. 5.3 depicts the four task clusters for Subject 1 (S1) and Subject (S2). It is clear that the clusters of T2 and T4 are close to each other, and away from both T1 and T3. This may indicate that the motor imagery (without movement) tasks are statistically closely related.



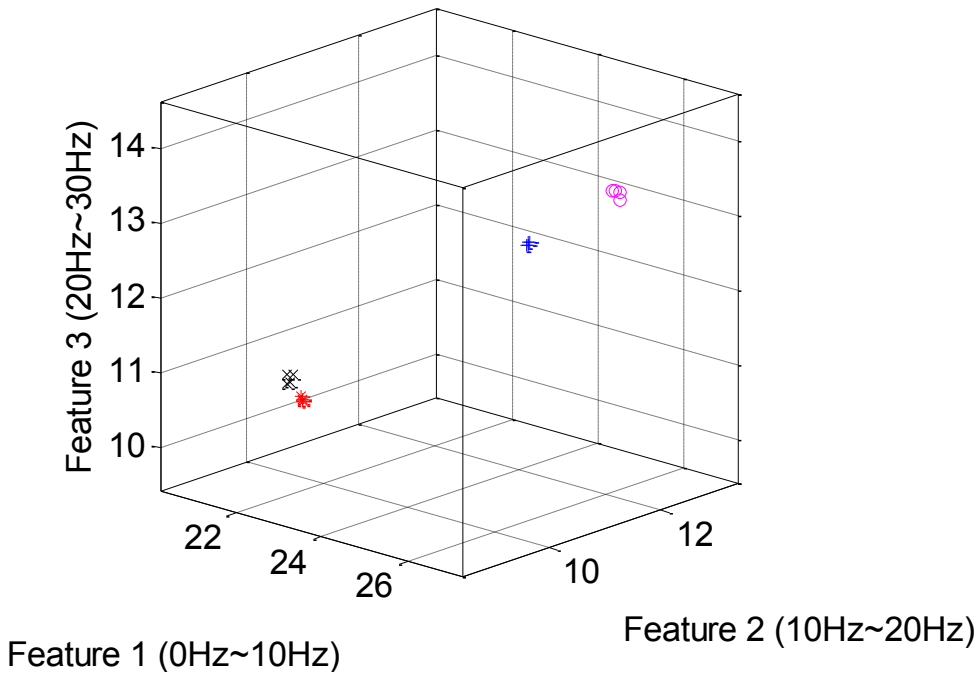


Fig. 5.3 Clusters of the first three dimensions of the feature vector (from four windows of 30 seconds duration, S1 and S2)

This work first investigates the biometric performance achieved when using EEG signals from different scalp regions. Nine selected electrodes clustered in three

distinctive scalp regions were selected for analysis (AF_3, AFz and AF_4 in the frontal lobe (F); C_1, Cz and C_2 in the motor cortex (M); O_1, Oz and O_2 in the occipital lobe (O). The positioning of the sensors is illustrated in Fig. 5.4. These regions were chosen to cover the anatomically significant areas of the brain involved in motor/imagery tasks [147], and to investigate the impact from other less effective region on EEG biometric performance [148].

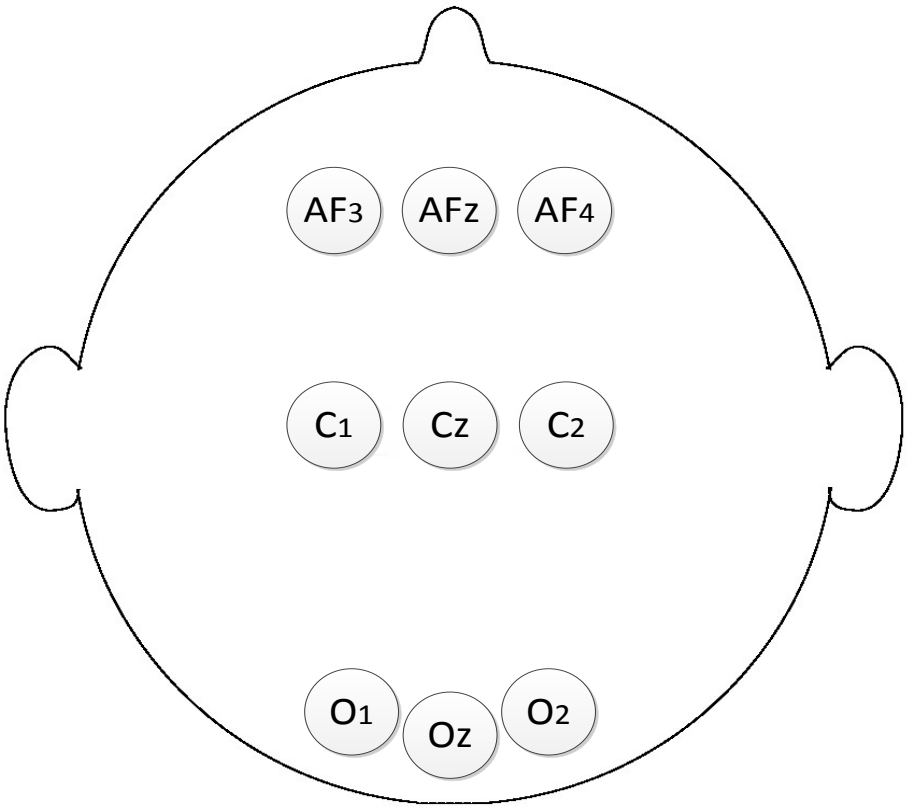


Fig. 5.4 Chosen Electrode Positions

Three experimental protocols are proposed to investigate several factors, which may impact the system's performance and thus address the research questions raised in the previous section. The three protocols were tested in both identification and verification scenarios. The abbreviations used above are combined, using the convention Task-Recording-Region to generate codes for the data subsets used in the experiments detailed below: e.g. TbEOF meaning Baseline task with open eyes and data from the frontal region electrodes and T1R1F refers to data from Task 1, Run 1 and frontal region electrodes.

5.4.1 Protocol P1

The goal of the first protocol (P1) is to investigate the impact of the sensitive regions used for data capture and the type of task on performance. Experiments performed using this protocol will also serve as a preliminary investigation to find the tasks with greatest biometric potential to be investigated further. The data subsets identified in P1 make it possible to explore the performance in each electrode region separately. The data from R1 together with R3 are randomly chosen as the training data and R2's data is employed for testing. P1 also identifies four groups of data subsets matching the four types of motor movement/imagery tasks (T1-T4) in MM/I to facilitate experiments to see the relative performance of each task type. The training and test datasets for P1 are shown in Table 5.1.

Table 5.1 Protocol 1

Data Set	Training Set	Test Set
Task 1	T1R1F+T1R3F	T1R2F
	T1R1M+T1R3M	T1R2M
	T1R1O+T1R3O	T1R2O
Task 2	T2R1F+T2R3F	T2R2F
	T2R1M+T2R3M	T2R2M
	T2R1O+T2R3O	T2R2O
Task 3	T3R1F+T3R3F	T3R2F
	T3R1M+T3R3M	T3R2M
	T3R1O+T3R3O	T3R2O
Task 4	T4R1F+T4R3F	T4R2F
	T4R1M+T4R3M	T4R2M
	T4R1O+T4R3O	T4R2O

5.4.2 Protocol P2

The purpose of the second protocol (P2) is to investigate the impact of using different mental/imagery tasks for training and testing of the system – the test data may be taken from a different task type to that used for training the system (all of the selected

nine electrodes are involved). P2 makes it possible to see if a mismatch between the training and testing task types can significantly impact performance. As in P1 the data from R1 together with R3 are used as the training data and R2's data is employed for testing. Additionally the data from the two baseline datasets are also used in this protocol for testing as illustrated in Fig. 5.5. In this figure each arrow signifies a combination of a training subset and a test subset that could be used in experiments.

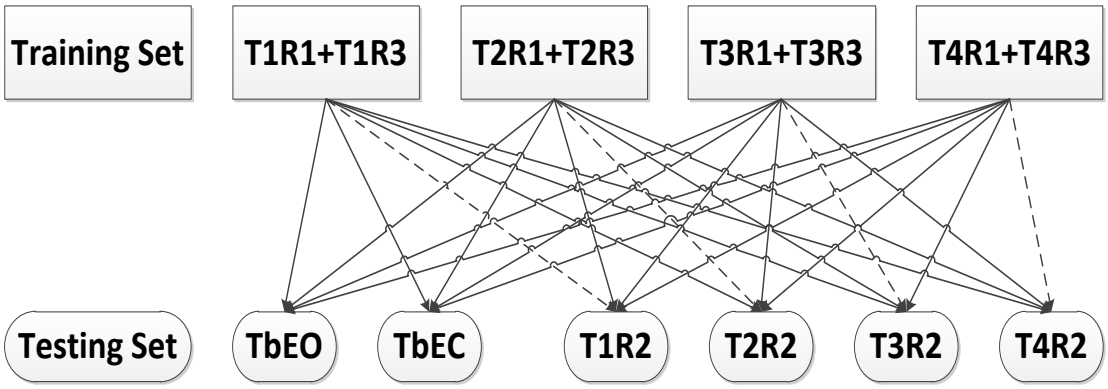


Fig. 5.5 Protocol 2

5.4.3 Protocol P3

The third protocol (P3) is designed to explore if data from different task types may be combined for the training the system to achieve better performance results. Test data from just one task type and recoding (T1R2) was used in this protocol. The training data was generated by combining (concatenating) an increasing quantity of data from different task type. The data subsets used in P3 for training and testing are shown Fig. 5.6.

Fig. 5.6 Protocol 3

5.5 Experimental Analysis

This section is devoted to the analysis of the results according to the three protocols proposed in the previous section. The experiments were conducted for both the identification and verification scenarios. In the identification the system will compare the query with all of the templates in the database. In the verification scenario, the subject claims an identity this is compared with the particular template which belongs to the claimed identity stored in the database. Therefore, there are two basic indicators to evaluate the system's performance in the verification scenario: the False Acceptance Rate (FAR), and False Rejection Rate (FRR), from which the Equal Error Rate (EER) may be computed [2]. One of the standard evaluation methods to combine these two indicators is the so-called Detection Error Trade-off (DET) curve, which will be used throughout this work to evaluate the proposed system [149]. The software that employed to generate DET curve was provided by the National Institute of Standards and Technology (NIST) [150].

5.5.1 Identification Scenario

The system's performance in the identification scenario will be investigated in this sub-section. Three aspects will be analysed according to the proposed protocols: 1)

the impact of electrode positioning on the recognition accuracy, 2) the influence from different tasks and 3) the impact of combining data from different tasks for training.

1) Test according to P1

Fig. 5.7 illustrates the impact of the electrode positions for different tasks. The results were generated by randomly selecting 75% of the feature data from R1and R3 to train the system, and 25% of the feature data from R2 to test it. The tests were repeated 100 times for generating the box plots. It is observed from the graphs that for different mental task, although the electrodes are clustered in three distinctive regions (frontal, motor and occipital lobes), there is no concrete evidence to indicate that any particular region could be a better placement for biometrics identification: for Task 1 and Task 4, the data from occipital lobe seems to provide relatively better performance. For Task 2, which is one of the motor imagery tasks, the motor cortex indeed provides slightly better performance. For Task 3 the frontal lobe provides the highest median accuracy.

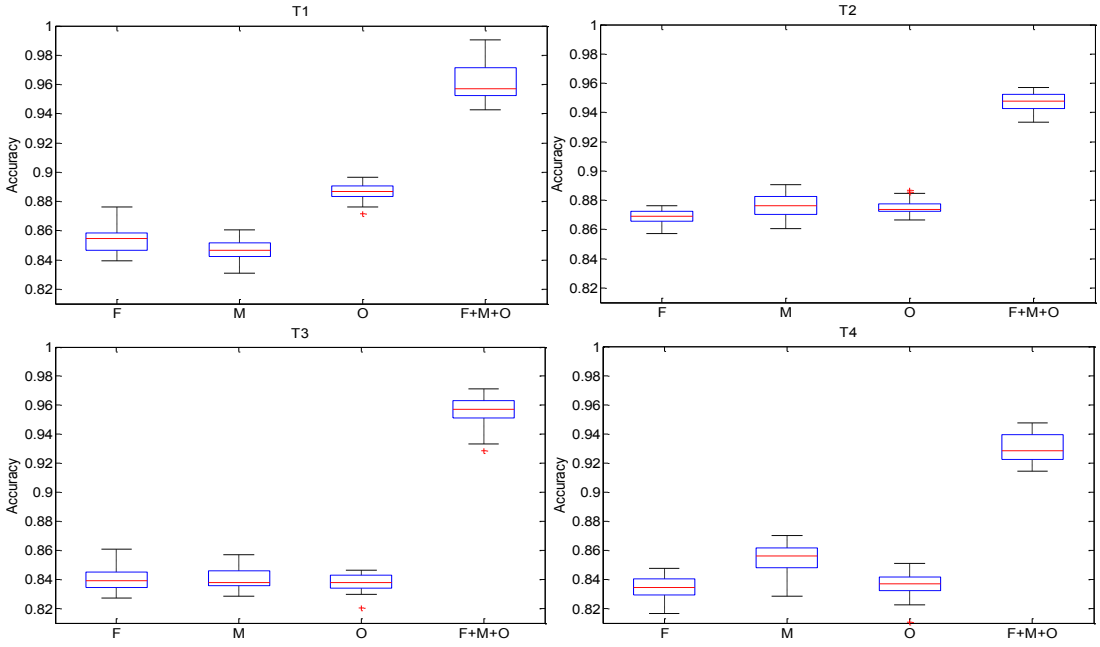


Fig. 5.7 The identification performance across four mental tasks for the three regions (three electrodes per region)

Further inspecting the identification accuracies of the four mental tasks, it seems that introducing feet movements may not be a good idea for biometrics recognition: the overall performance of T3 and T4 (movement or imagery movement of both fist and

feet) are both worse than T1 and T2 (movement or imagery movement of only fist). T1 seems to provide the best overall performance (in terms of both median accuracy and the variance of the performances): such evidence may indicate that the actual motor movement is a better choice over the motor imagery movement for biometrics scenarios. Based on these observations, the P3 protocol is focused on exploring the performance that can be achieved with T1.

In summary the results show that while the position of the sensors do not make a substantial difference in identification performance, the choice of the motor mental task of opening and closing the fists seems to outperform the other tasks in the database.

2) Test According to P2

For these identification scenario tests, the data from all the nine electrodes are combined through concatenation to evaluate the impact of the different (non-matching) tasks used for training and testing.

As in P1, the second run data (R2) of each task is used for testing and the data from the first and the third runs (R1 and R3, concatenated) are used for training the classifier. The two baseline recordings (EO and EC) were also used for testing to establish the usability of such data in a biometric context.

For TbEO and TbEC instead of 25%, 50% of the feature data were randomly selected for testing due to their shorter recording length compared to the other four tasks. The results indicate that the mental (non-resting) tasks used for testing do not make a significant difference in the results as indicated by the means of identification accuracy. By training with data from T1 to T4 separately and testing across the four mental tasks accordingly, the SDs of the four mean identification accuracies range from 0.0176 (trained by T1R1+T1R3) to 0.0378 (trained by T3R1+T3R3) indicating the stability of the performance results. Similar to the results obtained from P1 in the previous section, training with the data from T1 again has shown the highest and the most robust performance amongst the four mental tasks (Table 5.2).

Table 5.2 Results according to P2, across the four tasks

Test Set \ Training Set	T1R1 + T1R3	T2R1 + T2R3	T3R1 + T3R3	T4R1 + T4R3
T1R2	96.15%	91.22%	89.73%	89.49%
T2R2	96.44%	94.72%	86.48%	91.45%
T3R2	92.78%	90.13%	95.50%	88.42%
T4R2	95.01%	94.91%	87.12%	93.10%
TbEO	1.92%	1.43%	2.32%	2.08%
TbEC	3.45%	1.45%	1.92%	2.44%

The results suggest that when the training data was tested by the data from the same task, the variance of the CRR is usually smaller than when it is tested by data from different tasks. The median accuracy of non-matching training and testing sets, however, can be slightly higher (e.g. a system trained by T1R1 together with T1R3, tested by T2R2 provides better average performance). In short, the results suggests that given a particular type of task or tasks used for preparing the training data, the system may still be able to give acceptable results when an altogether different task type is used for testing. This allows for more flexibility from the perspective of both system designers and subjects in real-life biometric applications.

Interestingly enough, the performances observed while the mental activity tasks tested with the two baseline tasks (EO and EC) were much worse than they were tested with one of the movement/imagery tasks. One possible reason for such a significant difference in identification rate might stem from the energy difference of the signals, as the resting state seems to indicate the brain should be less active. However, our further investigation showed the signal energy (square of the signal amplitude in time domain) from these two kinds of state (resting and mental activity) has no considerable difference. This might also suggest the ambiguous nature of resting state: it is hard to establish whether the subjects are really in a resting state or not by simply measuring the signal energy.

3) Test According to P3

The results of the previous experiments have shown that the matching between the training and testing task types is not crucial for achieving good performance provided a non-resting task is used for training and testing. In this set of experiments, therefore, data from multiple task types are combined for training in the hope of achieving better classifier training. Separate experiments were performed with data from each individual run of Task 1 for testing and the best results, which are reported in this paper, where obtained when R2 was used for testing.

As it is shown in Fig. 5.8, as the size of the training set is gradually increased (data from all nine electrodes from the three scalp regions), the CRR also increases despite the training data coming from different types of mental activity. The rightmost column shows that when all the available data is used for training, the overall performance (in terms of the median accuracy and variance) drops slightly. These results suggest that adding more data may not always lead to improved performance due to the presence of redundancies and noise.

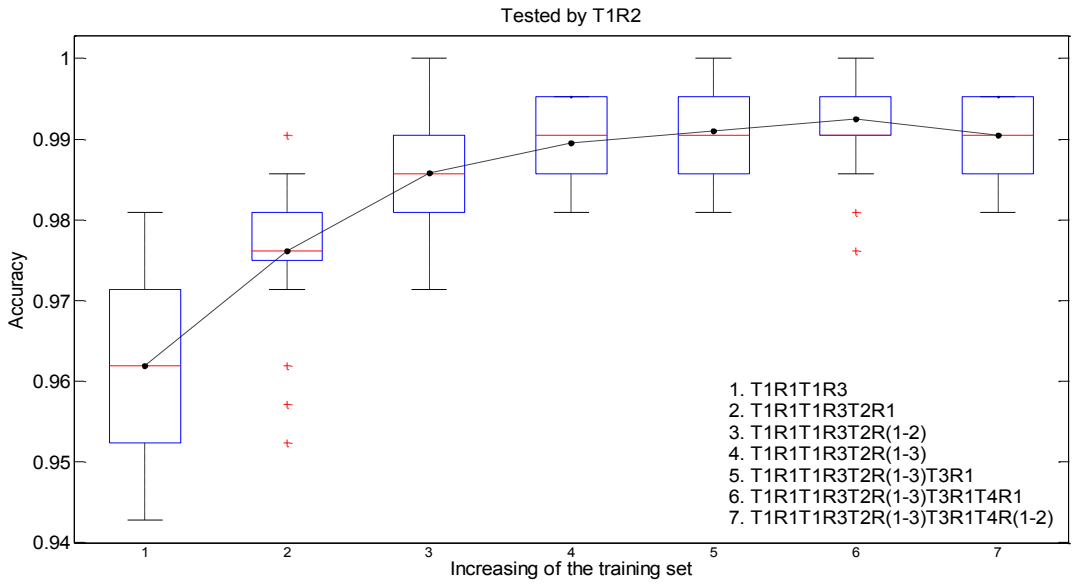


Fig. 5.8 Results according to P3

5.5.2 Verification Scenario

Performance of the system in the verification scenario is investigated in this section. The following three sub-sections are devoted to the experiments based on the

proposed three protocols. According to the results obtained from the identification scenario, some of the testing schemes explored in the previous sections are simplified.

1) P1: Analysing the impacts of different electrode positioning

The DET curves in Fig. 5.9 depict the performance of signals captured from three regions of the scalp. Data from different tasks was used for training and test but here only the results from T1 are reported as an example to support the arguments of P1 (results obtained from other tasks also shown similar trend). The data of R1 and R3 together were used for training the classifier and the data of R2 was used for testing the system's performance.

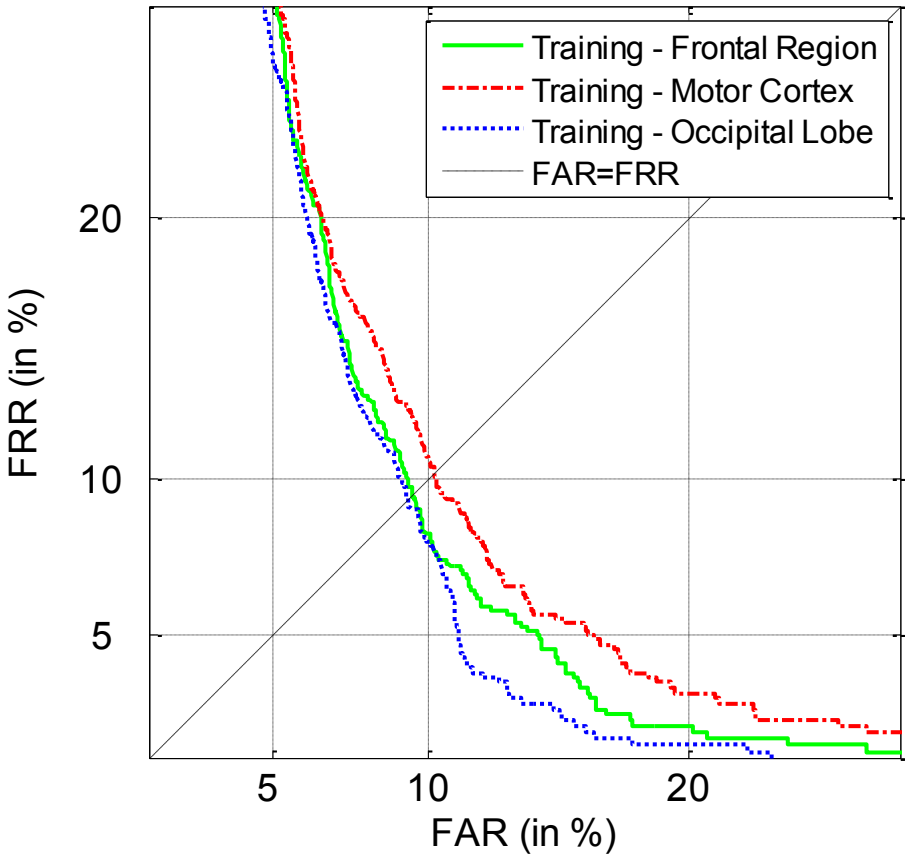


Fig. 5.9 T1 according to P1 in verification scenario

In Fig. 5.9 it can be observed that the occipital electrodes provided the best the overall performance amongst the three regions while the electrodes from the motor cortex seem to have the worst overall performance. The data obtained from occipital lobe and frontal lobe provided close EERs; data from motor cortex resulted clearly

higher EER. Though the results of Fig. 5.9 indicate the occipital lobe provided slightly better performance amongst the three major scalp regions, it is not clear that the impact from the electrode locations is significant.

2) P2a: The impact of mismatched training and testing tasks

The DET curves in Fig. 5.10 depict the performance when different types of task were used for testing the system, while it was trained by the data from Task1. Based on the results of P1 in the verification tests, the data provided by three occipital lobe electrodes were used here in P2a. The data of T1R1O together with T1R3O of these electrodes were used for training, the data from the middle runs of the four motor movement/imagery tasks were used interchangeably for testing. Additionally, the data from the two baseline tasks (TbEO and TbEC) were also used for testing.

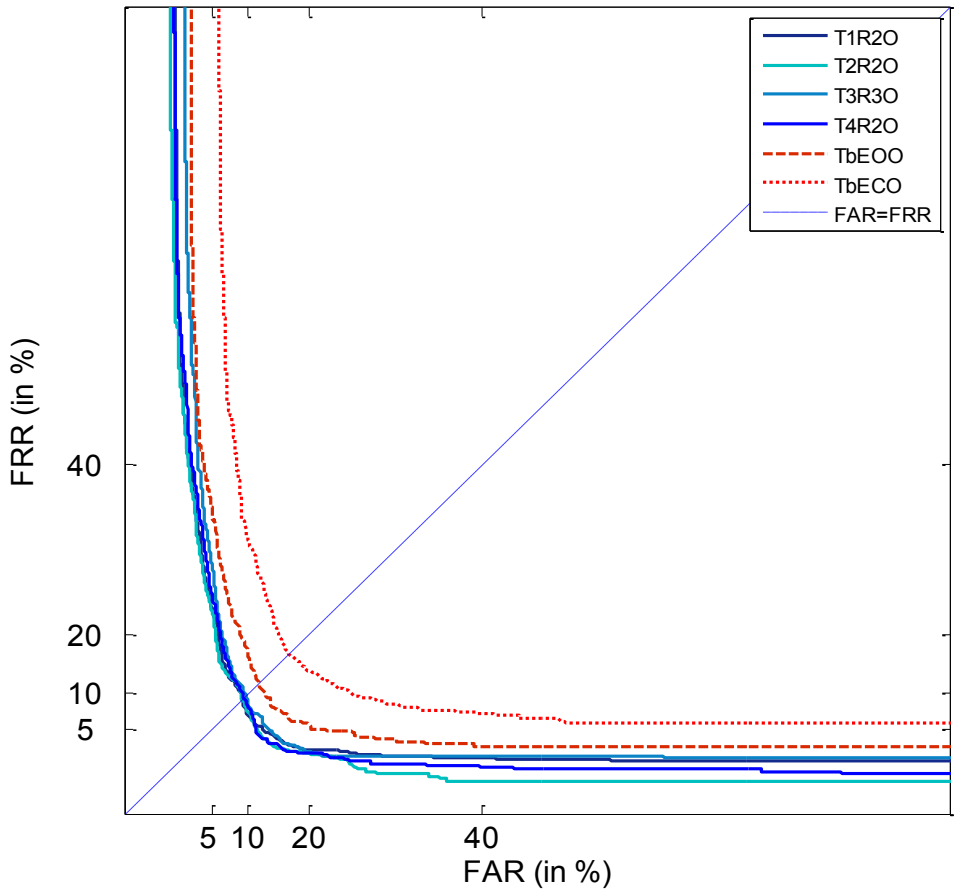


Fig. 5.10 Training with T1R1O+T1R3O, tested with different types of tasks, three occipital electrodes

Using the occipital lobe data of T1R2 for testing, i.e. training and testing with the same type of mental activity task, the equal error rate was 8.26%. Using the same

training set, when it was tested by the data from one of the other motor movement/imagery tasks, the performance did not vary much: EERs were 8.09% by T2R2O and 7.83% by T4R2O, respectively, which were even lower than when it was tested by the same type task data (T1R2O). Only while it was tested with T3R2O the EER increased to 8.93%, slightly higher than the result obtained from using T1R2O. The rest two curves which represent the results obtained from the data of the two baseline tasks, however, revealed much worse performances. Furthermore, the TbEOO curve indicates the performance when the system was tested by the eyes open baseline data, which is much better than the result provided by the data obtained while eyes were closed (TbECO).

By observing the results of Fig. 5.10, it seems that when the system was trained by the data from one motor movement/imagery task and tested by any of the other three different task data, the performance does not necessarily become worse than the task-matching test; when the movement/imagery task data was tested by the data generated during resting state, the performance degraded a lot. However, eye blinks in this particular circumstance does not seem to be a negative factor, due to the obviously better performance during the eyes open state compared with the eyes closed state.

3) P2b: Increasing the number of electrodes

The previously presented results suggest that there might be no obvious electrode positioning bias in the verification scenario (Fig. 5.9). The impact of different types of motor movement/imagery activities on the verification performance seems to be not quite substantial either, though the results of task-mismatched experiments indicate the data from different brain state (resting state) indeed significantly affect the recognition accuracy (Fig. 5.10). Furthermore, the performance achieved using only three electrodes may not be sufficient enough for state-of-the-art biometric applications (Fig. 5.9 and Fig. 5.10). Therefore, two different approaches were used to improve the verification performance: (1) Increasing the size of training data by using more electrodes; (2) Increasing the size of training set by combining data from different task types (according to P3). The utilization of the second strategy was based on the experimental results shown in Fig. 5.11.

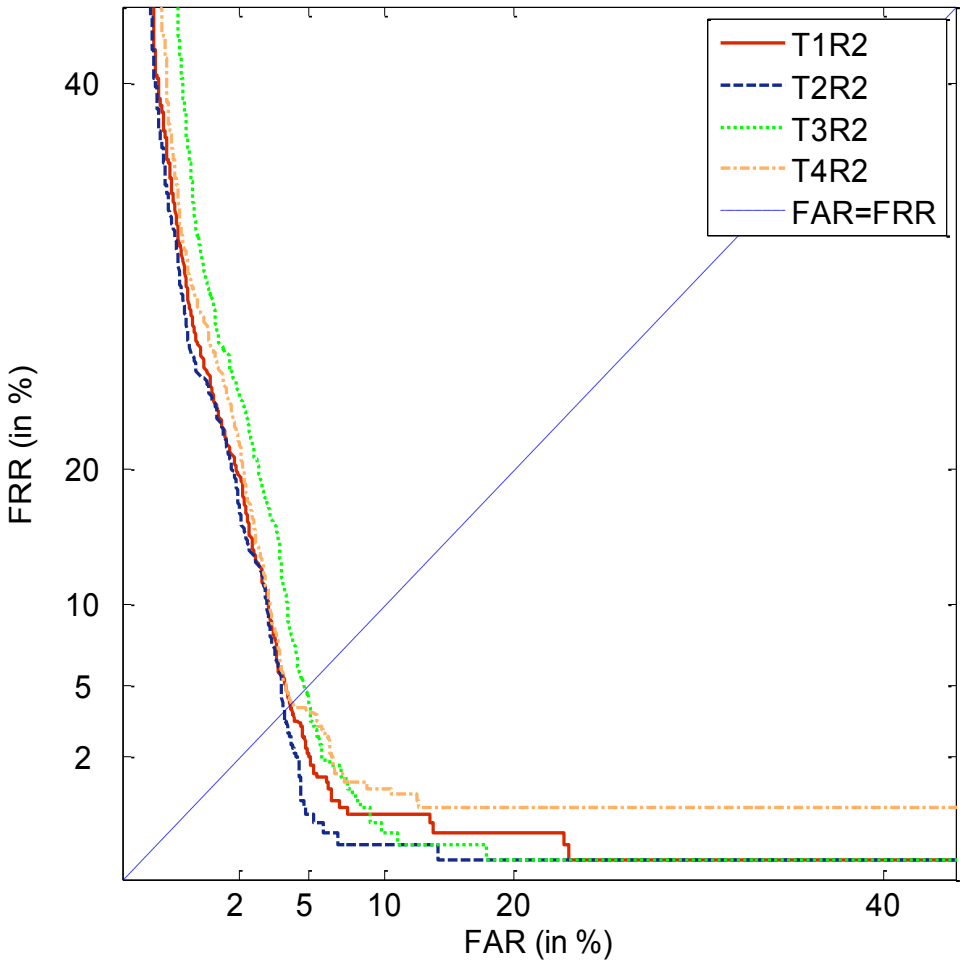


Fig. 5.11 Training with T1R1+T1R3, tested with R2 from different types of motor movement/imagery tasks, nine electrodes from the three regions

For the first approach, data from all the nine electrodes (in three regions) were employed. Trained by T1R1 together with T1R3, a series of tests were conducted to measure the improvement of the performance and the results are shown in Fig. 5.11. Similar to P2a, the data from the middle runs (R2) of the four motor movement/imagery tasks were used for testing. It can be observed that when the system was trained with the data from T1 and tested by T2R2, the EER gained the lowest value of 2.785%, which is even lower than when it was tested with the data of the same task T1R2 (6.9% of EER). The best performance in this training-testing scheme has improved almost three times, by increasing the number of employed electrodes from three to nine ((8.26%) / (2.785%) $\approx$ 2.97).

4) P3: Concatenating different task data for training

The data from different motor movement/imagery tasks were concatenated for improving the training of the system and increasing performance. The data from all nine electrodes were employed for training the verification system.

A series of DET curves in Fig. 5.12 depict the results obtained from different training data sets used in P3. Data from T1R2 alone was used as the test set; the data used for training the system were increased gradually. As the data of T1 and T2 (motor movement tasks) were found to have provided slightly better performance than T3 and T4 (motor imagery tasks), these were combined to train the system. The system achieved the lowest EER of 2.87% by training with the largest amount of data (combining all the data of T1 and T2). The performance of using data from the single run for training were also explored, i.e. trained by the data of T1R1 and T1R3 separately: the results show that using only the data from a single run, the verification performances are much worse compared with the combination of runs and tasks. It may be reasonable to expect that if more data from different motor movement/imagery tasks were combined in such a way, the EER may be further reduced. This pattern has indeed been observed for the identification results in Fig. 5.8.

The DET curves of Fig. 5.12 could be divided into three groups. For the group with the smallest training set, using the data of single runs from a single task for training, the lowest performances were observed and the training data of R3 provide better results than R1. This performance variation may also indicate the instability of employing EEG in biometrics verification scenario while using single runs for system training. The performance improved with increasing the training data volume (from multiple data recordings) and alleviated the performance variation. The lowest EER of about 4.5% was achieved when all the data from T1 and T2 were used for training. It is worth mentioning that the inflection points observed on the DET curves become less distinct as the training data increases. The results indicate that a system with zero FRR and 7% FAR may be achievable which might be suitable for some application scenarios.

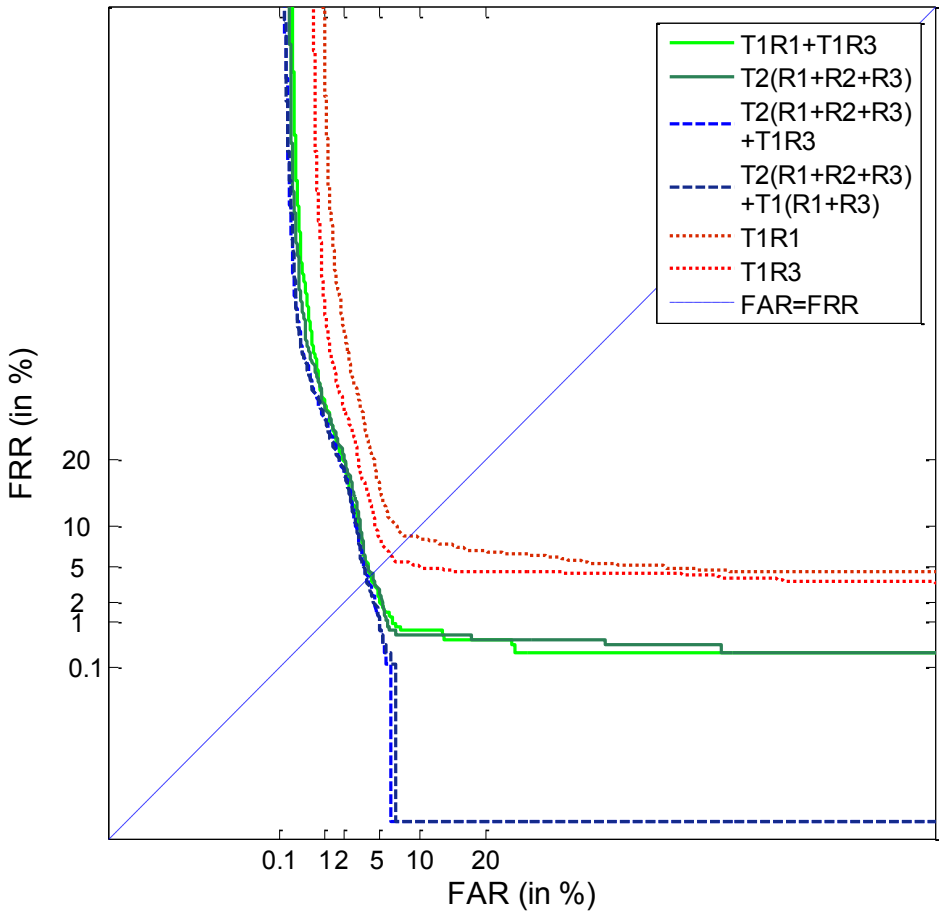


Fig. 5.12 Tested with T1R2, cross-task increasing the training data size

5.6 Summary and Discussion

In this section a brief summary of the key contributions in this chapter is provided, followed by a discussion of some important issues that were raised.

A. EEG Features

A new feature for EEG biometric signals based on the derivative of wavelet coefficients is proposed. Differentiation is a conventional function used to measure the changing rate of signal series. The motivation for calculating the differentiation of the wavelet coefficients is to capture the possible instantaneous change of the signal.

The use of multiple scales of wavelet coefficients is also investigated. By combining the coefficients from different scales, more information may be captured representing

both time (from the higher scales) and frequency (from the lower scales) characteristics of the signals.

B. Evaluations

Three protocols were constructed to verify the questions raised in Section 5.2, mainly aimed at establishing the impact of training strategies and data volume on performance.

The results of P1 indicate that for the proposed experimental design there is no clear difference amongst scalp regions.

P2 was designed to answer whether using different mental tasks for testing would impact performance. It can be observed from Table 5.2 that the sensitivity to different mental tasks used for testing was quite limited with the mean performance changing very little. However, when one of the resting tasks is used for testing there is a large drop in performance.

P3 is designed to investigate whether it is applicable to aggregate data obtained from different types of mental activity from separated recordings to build a more robust training al model. Considering the possible similarities between human voice and EEG signals, it is noted that similar strategies have been successfully implemented for text-independent speaker recognition systems [151]. Results from P3 clearly indicate that increasing the training data volume, irrespective of the type of mental activity used, improves identification and verification performance (Fig. 5.8 and Fig. 5.12).

5.7 Conclusions

In this work the impact of mental activity on the performance of an EEG-based biometric system using wavelet features has explored. The work has investigated the impact of different mental activities used for training and testing the system. The overall conclusion is that there appears to be substantial flexibility in the choice of mental activity used for training and testing such systems. The work has also indicated that data from different types of activity may be aggregated to provide more robust training of the system without any adverse effects. This flexibility with regards

to types of mental activity could result in systems that are easier to develop, deploy and use in a range of applications. Future work will be focused on evaluating the robustness of this approach when collecting data with long time intervals between training and testing as well as data from low-cost EEG sensors.

Chapter 6

Wavelet-based Features and HHT-based Features

The choice of features is essential in achieving good biometric recognition performance; many features have been proposed in the literatures (Chapter 2). The PSD and AR model coefficients have been amongst the most popular features in EEG biometrics since they were first introduced to this field about 15 years ago [5][6]. However, the conventional PSD features only reveal static content (spectral power) of the signals without considering the dynamic non-stationary characteristics of brainwave. Typical AR features, on the other hand, are time domain coefficients, which may be difficult to reveal the frequency property of the signals. In this chapter, four new and arguably better performing features are proposed: the theoretical arguments for their suitability are further supported by their respective experimental results.

This chapter contains five sections: two wavelet-based features are developed and explored in Section 6.1 and Section 6.2, respectively. In Section 6.3, two closely-related novel features based on Hilbert-Huang Transform (HHT) are described, and their experimental evaluation is presented. In Section 6.4, a comparison of the proposed features is given, followed by a discussion of comparative experimental results. The conclusion is presented in Section 6.5.

6.1 Time-derivative of Wavelet Coefficients

In the Preliminary Investigation chapter (Chapter 4), a series of different wavelet-based features were introduced and experimentally evaluated. The results indicated computing the standard deviation (SD) of wavelet coefficients for each segmented window was a good choice as feature for the MM/I dataset. In this section, an enhanced feature based on the notion of SD feature is developed.

In this investigation it is proposed to employ the Wavelet Packet Decomposition transform for feature extraction, which includes a full decomposition of the signals into multiple levels using both wavelet and scaling functions [59]. In order to maximize the use of both time and frequency properties of the signal, not only the coefficients from the lowest level but also the higher level coefficients were employed as the primary features. The EEG signals were decomposed up to Level 3 (Fig. 6.1). This allows the signal to be divided into eight non-overlapped wavelet bands. The nodes marked in brackets in Fig. 6.1 indicate the bands and levels used to construct the feature vector.

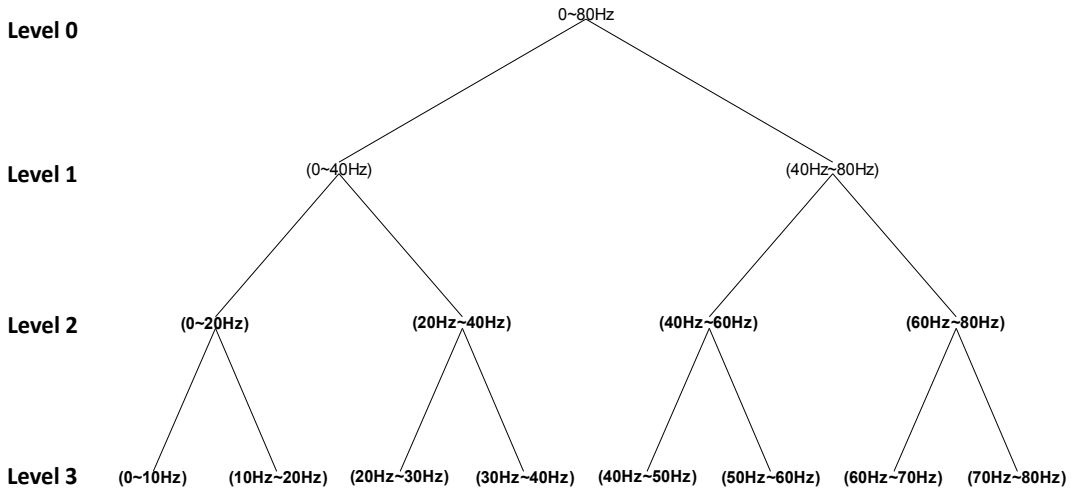


Fig. 6.1 Wavelet Packet Decomposition for the Proposed System

The original SD-based method employs the resulted wavelet coefficients of each window for the feature extraction directly, whereas the new strategy adds an extra step of computing the derivative of the wavelet coefficients in each of the selected bands before computing the SD. This operation is aimed at enhancing the temporal variations in the wavelet signal and was found to significantly improve the performance, which is similar to the notion of image gradient which represents the change in intensity or colour of an image [146]. Depending on the types of available datasets (MM/I dataset with four tasks) in the work, accuracy improvements of 5%~8% were found by computing the derivative of the wavelet coefficients compared with not employing this signal enhancement stage.

After calculating the derivative of the wavelet coefficients, the standard deviation of each wavelet band for each window is computed (*SD*(*Diff*)). The main motivation for this step is to reduce the dimensionality of the training data: each window with 4800 samples is represented by twelve SD values representing each of the wavelet bands.

The overall system diagram related to this method can be found in Chapter 5, Fig. 5.2. The effectiveness of the proposed time-derivative of wavelet coefficients feature is to be investigated and compared along with other proposed features in Section 6.4, in terms of their biometric recognition performances.

6.2 Wavelet-DCT Coefficient Feature

Feature extraction is one critical step in developing EEG biometric system. After more 15 years of research, most of the proposed systems are still based on extracting one or both of the following features for classification: Power Spectral Density (PSD) [42][44] and Autoregressive Model (AR) coefficients [55][99]. These conventional features, however, may not be able to capture enough biometric information given the data from only a small number of electrodes are available, particularly while without applying the feature fusion strategy. Reducing the number of employed electrodes is one of the main research trends in EEG biometric field. It is noticed from the existing literature that the performance of using conventional features extracted from small number of electrodes still need to be improved and the new features are actively developing.

Recently, Bai *et al*. [101] reported to use the Visual Evoke Potential (VEP) of EEG signals for person identification. A series of techniques, including Genetic Algorithm, Fisher Discriminant Ratio and Recursive Feature Elimination were employed to reduce the adopted electrodes for less-intrusive user experience. Data from 32 out of 64 electrodes were selected for testing in a self-collected database of 20 subjects. The best identification rate of 97.25% was achieved using a Support Vector Machine classifier.

Phung *et al*. [68] proposed to use Shannon Entropy (SE) as feature for fast EEG-based person identification. A database which contains 40 subjects was employed. EEG data

captured from 23 electrodes was employed for feature extraction. A comparison of using the conventional Autoregressive coefficients (AR) as features was presented: it was found that using SE was 2.3 to 2.6 times faster than the feature extraction based on using AR model to achieve comparable accuracy (97.1% for SE versus 97.2% for AR).

Gui *et al*. [45] proposed two methods for EEG feature extraction: Euclidean Distance (ED) and Dynamic Time Warping (DTW). The proposed methods were tested in a self-collected database with 30 subjects using a 74-channel EEG cap: only the data obtained from four electrodes (Pz, O1, O2, Oz) were used. It was found using ED method over 80% of accuracy was achieved whereas for the DTW method it was about 68%.

Same as the human voice signals, the EEG signal is considered as a nonstationary modality [152][153]. A preliminary investigation of extracting the conventional Mel-frequency Cepstral Coefficients (MFCC) for EEG person identification was conducted by Nguyen *et al*. [69]: for a population of 20 subjects (one subset of a public database with 122 subjects) the identification rate of 92.8% was achieved using the data from selected eight electrodes. It suggests MFCC feature which transferred directly from voice recognition is also effective in revealing the biometric information of EEG signals. MFCCs are commonly derived as follows [154][155]:

1. Take the Fourier Transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the Mel Scale [156], using triangular overlapping windows.
3. Take the logs of the powers at each of the Mel frequencies.
4. Take the discrete cosine transform of the list of Mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

In this work a new wavelet-based feature is developed, which shares certain same characteristics of the standard MFCCs extraction process. In fact, the notion of combing Wavelet Transform (WT) and Fourier Transform (FT) has been noticed and

implemented in the speaker recognition field [157][158]. However, all of the reported algorithms follow the order of computing *Fourier Transform-Mel Log Powers per Window-Wavelet Transform* for feature extraction. The proposed EEG related algorithm, on the other hand, performs the wavelet transform first, then the logarithm of the wavelet coefficients per window is computed, finally the Discrete Cosine Transform (DCT) is computed. The resulting coefficient with the highest energy is used as the feature for pattern classification. The overall flow chart of the algorithm is illustrated as follows:

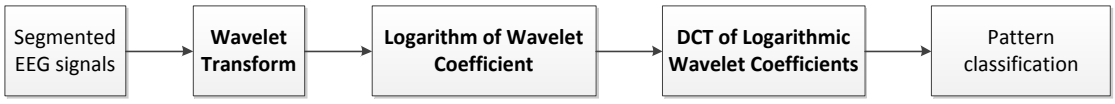


Fig. 6.2 Overall feature extraction process of Wavelet-DCT decomposition

As Fig. 6.2 indicates, the main difference between the proposed feature extraction method and the conventional MFCC algorithm is replacing the first step of Fourier Transform by the Wavelet Transform. The multi-scale decomposition characteristics of WT may be better distinguishing the biometric information buried in the time domain signals. The conventional MFCC algorithm employs the first 12-39 resulting coefficients after DCT as features for voice recognition [159]. In the proposed method, only the first/dominating resulting coefficient (per window) has been found effective and used as feature subsequently: all of the remaining DCT coefficients are found to make no contribution to recognition, according to the extensive experimental evaluations.

The main feature extraction process of the proposed method could be divided into three steps, highlighted by the bold letters shown in Fig. 6.2. The first step is performing the Wavelet Transform to each time-domain window after segmentation: here the Wavelet Packet Decomposition (WPD) is employed [59].

The second step is to compute the logarithm of the resulting coefficients for each post-windowed wavelet band. The motivation of performing this trick shares similar reason with the computing of MFCC features: the logarithm transforms the results to a non-linear scale, which serves as a magnifying operator in the proposed algorithm. For the

proposed method it is found the resulting logarithmic wavelet coefficients are more effective in person recognition, compared with the non-logarithmic features.

The logarithm of the wavelet coefficients for each window is further fed into a discrete cosine filters bank, result a series of Wavelet-DCT coefficients [146]. The final step is mathematically expressed by (6.1) as follows:

$$y(k) = w(k) \sum_{n=1}^{N} log\left[WT_{\psi}\{x\}(a,b)\right] \cdot cos\left(\frac{\pi}{2N}(2n-1)(k-1)\right), \tag{6.1}$$

where $w(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 1, \\ \sqrt{\frac{2}{N}}, & 2 \le k \le N. \end{cases}$ $k=1,2,\ldots,N.$

The distribution of the resulting wavelet-DCT coefficients shows in Fig. 6.2 is a visual illustration of the resulting coefficients after applied the proposed feature extraction algorithm. Preliminary evaluation results of DCT (after WT and logarithm) indicate the coefficient with dominating amplitude is the most discriminating feature. It is found usually the Delta (<4 Hz) and Theta (4 Hz-7 Hz) bands of the EEG signal always provide similar trend as the plot in Fig. 6.2, though for different individual the maximum value of Wavelet-DCT coefficients are different. It may indicate the user-distinctive information in the low frequency of EEG signals is dominated within a very narrow frequency range. Experimental results confirmed this conjecture: using other coefficients (instead of the dominating coefficient) as feature, the identification performance was less than 10% for a database of about 100 subjects. It is worth mentioning that such pattern as in Fig. 6.2 tends to happen in relatively low frequency of EEG signals (after wavelet transform), in high frequency (more than 60 Hz) such phenomenon becomes let obviously, i.e. the first coefficient in high frequency becomes less effective in biometric recognition.

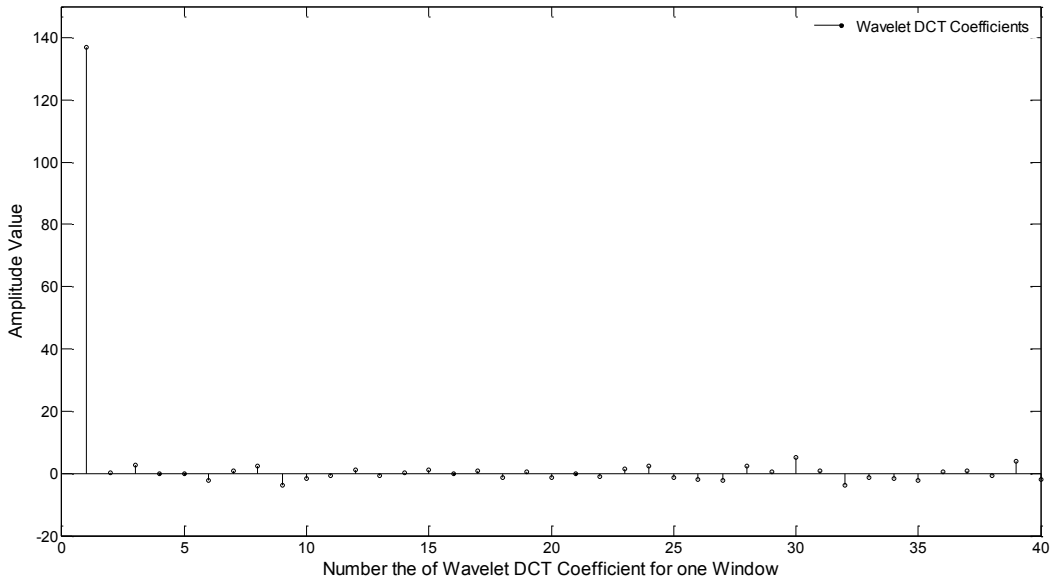


Fig. 6.3 Typical Wavelet-DCT coefficients per window, the illustrating plot obtained from 0~10 Hz of wavelet coefficients (generated using MM/I dataset)

One of the reasons that for each wavelet band/sub-band only the first coefficient is found effective may be also related to the results of Wavelet Transform. The segmented signals after WT become a series of wavelet coefficients in the "wavelet domain", these coefficients generally corresponding to a certain frequency range in the conventional frequency domain. However, frequency characteristics of the wavelet coefficients within each wavelet band are not equally distributed over the entire frequency range but clustered in a relatively narrow range. For this reason, the logarithmic DCT may further magnified the dominating frequency and lead to very small number of useful Wavelet-DCT coefficients (often just one dominating coefficient is effective).

To summary the proposed method: the raw signals were segmented into multiple time domain windows first, for each window the wavelet packet decomposition (WPD) was performed [136]: the signals were decomposed into four levels (level 0 to 3), only the wavelet coefficients between 0 to 60 Hz were preserved for further feature extraction. Therefore, 9 overlapped wavelet bands were kept for each time domain segment. Based on the illustrating steps of Fig. 6.2, the logarithm of the resulting wavelet coefficients was computed. DCT is then computed to generate the informative features for classification. It needs to mention that only the real part of the resulting DCT pair

is preserved. The resulting Wavelet-DCT coefficients for one wavelet band are shown in Fig. 6.3. Finally, only the coefficient with dominating amplitude (energy) is fed for classification. The features are trained by a simple 1-nearest neighbour classifier [82].

In order to test the performance of the proposed feature extraction algorithm, it is important to compare it with the related features in the literatures. Conventionally, after FT the number of resulting coefficients is too large to use directly, as it will hugely increase the dimensionality of the feature vector. High dimensional feature may lead to the "curse of dimensionality", especially when instance-based classifier such as *k*-NN is employed.

Many operations were employed to reduce the vector length. For example, Phung *et al*. [68] proposed to employ Shannon Entropy as feature. Abdullah *et al*. [61] proposed to compute mean and standard deviation as features to achieve the dimension reduction purpose. Gupta *et al*. [60] proposed to using the energy of resulting wavelet coefficient as feature, which can be obtained by computing the variance of the coefficients. In this study, some of these related features are also computed to compare their performances with the proposed Wavelet-DCT feature.

Two popular databases introduced in Chapter 3 were employed for this investigation: "UCI EEG Database Dataset" [111] and "EEG Motor Movement/Imagery Dataset" [15].

For both of the databases, only the data from the Cz (centre of the scalp) was employed for person recognition in this study. The preliminary investigations (in Chapter 4) indicate the EEG data obtained from Cz location tend to provide the best and most robust recognition performance.

6.2.1 Experimental Analysis

A series of investigations using the wavelet-DCT feature are presented in this section. The section is divided into three sub-sections. Section *6.2.1.1* is devoted to analysing identification sensitivity of the proposed feature using Cumulative Matching Curve (CMC), followed by the verification performance of the proposed feature and the

relevant wavelet features reported in Section *6.2.1.2*. The comparison with other related features are introduced in Section *6.2.1.3*...

6.1.1.1 Identification Sensitivity

The proposed wavelet-DCT feature is derived from time domain segments. The length of the segmented window is an influential factor in designing the EEG-based biometric system, using bigger window for feature extraction tend to achieve higher recognition rate. The window size also affects the user experience: the bigger the window size, the longer minimum time from the user is needed for testing/query.

The identification accuracy of the proposed system is shown in Table 6.1, using the UCI VEP database. A series of identification tests have been conducted: for the proposed feature, from around 10 seconds to about 30 seconds per window. The reported results are obtained by averaging 50 times of system training and tests: each time 70% of the available features (60 seconds' recording) is randomly picked for training and the rest used to test the system performance.

Table 6.1 Impact of window size (samples/window) for feature extraction: using UCI VEP database

Samples/window (Second)	2816 (11s)	3520 (13.75s)	4224 (16.5s)	5280 (20.63s)	7040 (27.5s)	8448 (33s)
Accuracy	77.86%	**90.20%**	77.31%	80.21%	**96.46%**	84.22%

Several features introduced in the previous sections are employed to compare with the proposed wavelet-DCT feature in an identification scenario. Fig. 6.4 shows the CMCs obtained by employing those features, using the UCI VEP database. It is clearly shown that the proposed feature provided the highest rank 1 identification rate. The Max-Min feature, which is derived by computing the maximum wavelet coefficient minus the minimum wavelet coefficients, provided the second best performance of about 64%. It is found SD feature (proposed by [61]), derived by computing the standard deviation of the resulting wavelet coefficients quickly surpassed the Max-Min feature after rank 3. It is interesting to notice though the computation processes of SD and variance are closely related, their identification performances are found rather different.

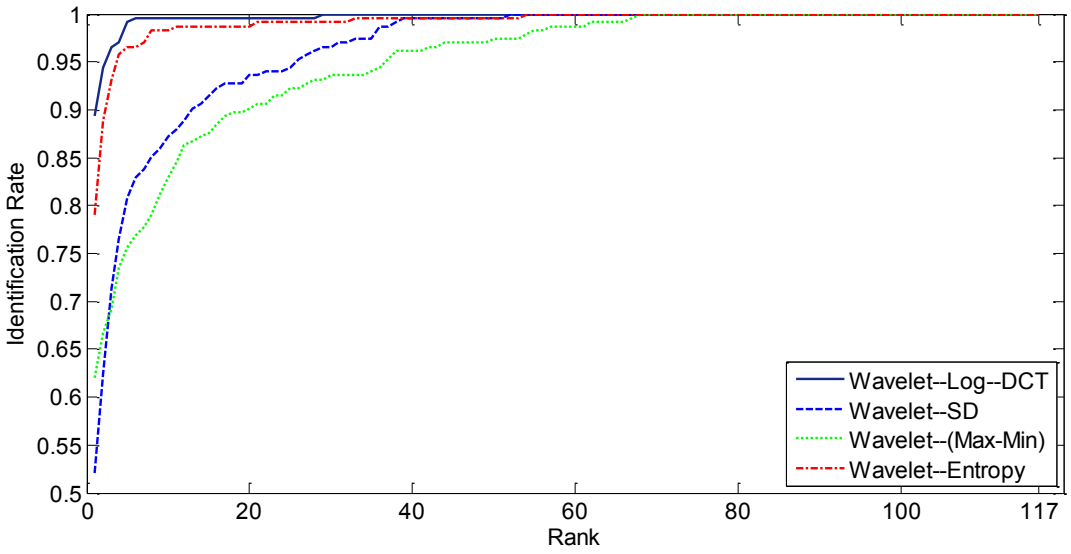


Fig. 6.4 CMCs of several features, using 117 subjects from UCI VEP database

The identification tests using the MM/I dataset also have been investigated. The CMCs shown in Fig. 6.5 indicates a much better performance than using the data from UCI VEP database (99% versus 90%). Considering the same recording length which purposely kept for comparison (60 seconds) and the same electrode location analysed (Cz), the results in Fig. 6.5 may suggest: 1) MM/I dataset is better noise-removed or 2) VEP stimulus is less effective than resting state with eyes close for EEG biometrics. Similar pattern is found in MM/I dataset: the newly proposed Wavelet-DCT feature provided the highest rank 1 performance; the SD feature (proposed by [61]) provided the second best rank 1 identification rate; same as it is for VEP database, the Variance feature (proposed by [60]) reveals comparable performance with feature derived using Shannon Entropy (proposed [68]). However, it is found from both Fig. 6.4 and Fig. 6.5 that Mean feature provided the worst identification performance.

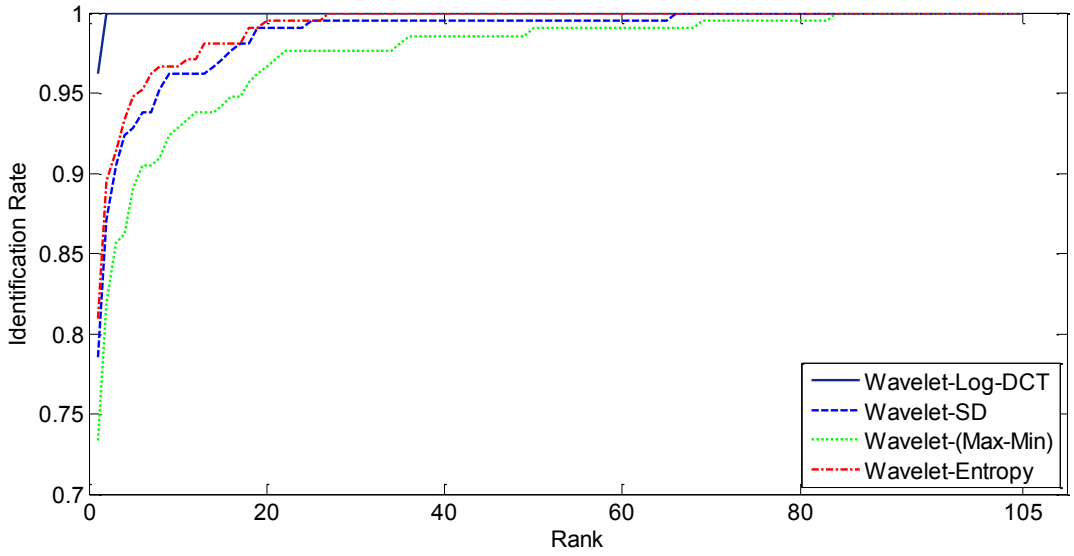


Fig. 6.5 CMCs of several features, using 105 subjects from MM/I dataset

It should be noted again that, the results in Fig. 6.4 and Fig. 6.5 are obtained using the segmented recording of 13.75 seconds per window, it is predicted that with the window size increasing, the identification performance could be further improved. The next section is devoted to the investigation of the performance of the proposed feature in the verification scenario.

6.1.1.2 Verification Sensitivity

The proposed wavelet-DCT feature is also tested using UCI VEP database and MM/I dataset in a verification scenario. The publicly available curve plotting package contributed by National Institute of Standards and Technology (NIST) [150] is employed to generate the Detection error trade-off (DET) curves.

Fig. 6.6 shows the resulting DET curves using UCI VEP database. For comparison purposes, the plots include the proposed feature and the related features. Wavelet-DCT feature here again shows the best verification performance: the Equal Error Rate (EER) between 2%~5% is achieved, the second lowest EER of more than 10% is provided by SD feature.

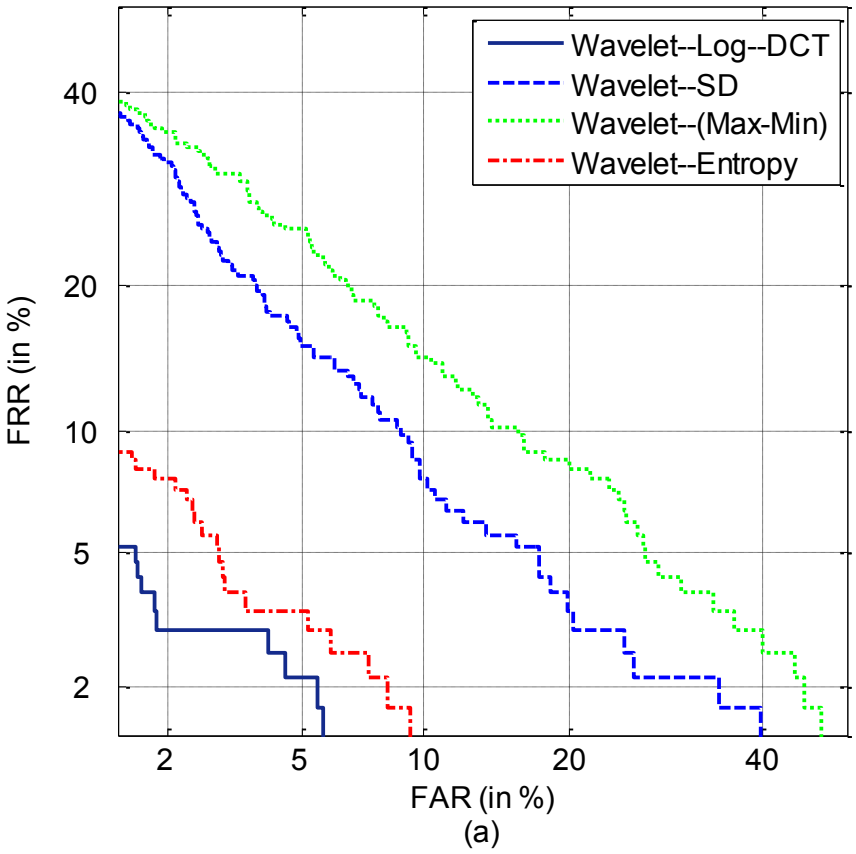


Fig. 6.6 DET curves of several features, using 117 subjects from UCI VEP database

To investigate the biometric recognition stability of the proposed feature for different database, the verification performance of the MM/I dataset is again used for comparison. Fig. 6.7 indicates the proposed feature provided about 1% of EER and the SD feature achieved the second best performance with EER of about 6%.

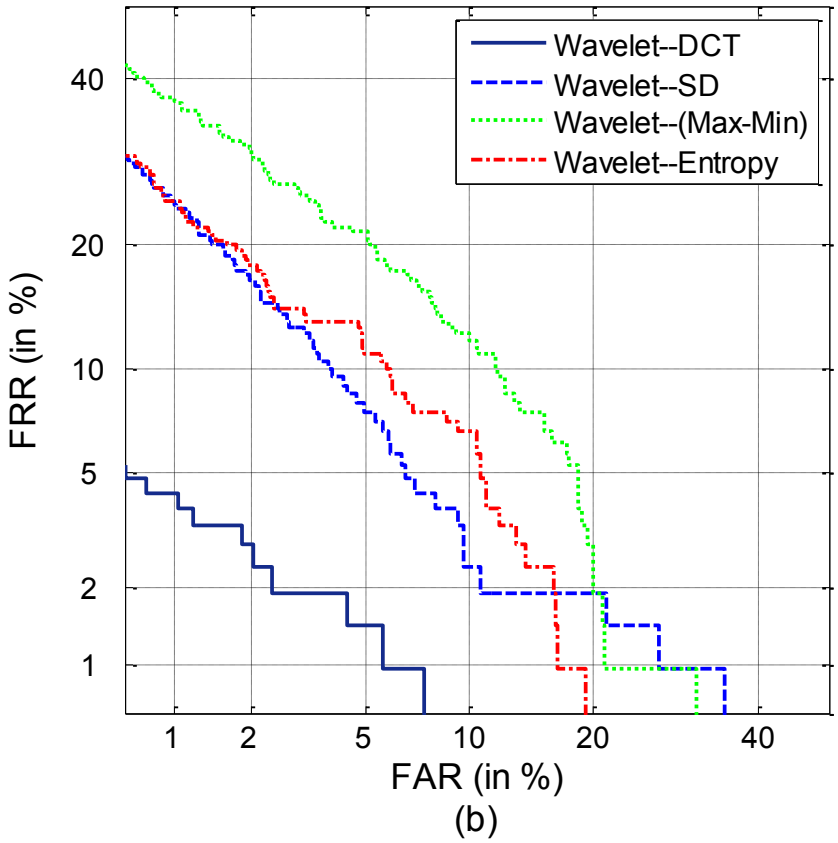


Fig. 6.7 DET curves of several features, using 105 subjects from MM/I dataset

Similar patterns can be found between Fig. 6.6 and Fig. 6.7: for both databases the proposed Wavelet-DCT feature provided the best performance and the Mean feature provided the worst verification rate. It is clear that the proposed feature revealed both good and stable verification performance. Considering the equal recording length for analysis (both 1 minute) and the same electrode employed (Cz), it seems the MM/I dataset may possess higher quality data than the VEP database for biometric recognition.

6.1.1.3 Comparative Analysis

In this section, the proposed feature is compared with other non-wavelet-based features. All the results reported here are based on UCI VEP database and the MM/I dataset.

Table 6.2 shows the results obtained by different features while using UCI VEP database. The PSD and AR coefficient features are conventional features: high identification rates were achieved using them. However, it is shown in the literature

that large number of electrodes has to be employed to maintain such high recognition rates. Though provided slightly lower recognition rate, the proposed feature extraction method with only single electrode offers much less intrusive Brain Computer Interaction (BCI). It also should notice that the result listed in Table 7.2 using Wavelet-DCT feature was obtained using feature extracted from window size of 13.75 seconds, the identification rate could reach 96.46% while the EEG recording were segmented into 27.5 seconds per window for feature extraction (see Table 7.1).

Table 6.2 Feature sensitivity for identification scenario, UCI VEP database

Features	PSD & AR coefficients [74]	Univariate AR model [56]	Root Mean Square [66]	Wavelet-DCT
Subjects	20	120	116	117
Electrode(s)	61	64	64	1 (Cz)
Accuracy	100%	98.96%	95.1%	90.20%

Table 6.3 shows the comparative results from employing MM/I dataset based on using different features. Similar pattern is found from MM/I dataset: using one feature for classification (with fusion of multiple features), the best identification rate is obtained by using the proposed feature from single electrode.

Table 6.3 Feature sensitivity for identification scenario, MM/I Eyes Open subset

Features	Spectral Coherence Connectivity [31]	Power Spectral Density [31]	Eigenvector Centrality [32]	Wavelet-DCT
Subjects	108	108	109	105
Electrode(s)	56	56	64	1 (Cz)
Accuracy	75.86%	86.91%	96.9%	98.6%

The main advantage of the proposed feature extraction algorithm is the small number of employed electrode for person identification. The existence of biometric information in EEG signals has been justified by many researchers, but for the sake of user-friendly experience, using single electrode for person recognition is one important research trend.

In this study a novel feature extraction method is proposed, the dominating Wavelet-DCT coefficient is found the most effective in person recognition. Only the electrode in the Cz location has been employed for data analysis. Both the identification and verification scenarios were explored, the results compared with multiple related

features appeared in the literatures. The comparative analysis was conducted to objectively illustrate results reported by other researchers using the same databases. In the next section, two novel features based on Hilbert-Huang Transform are proposed.

6.3 Features based on Hilbert-Huang Transform

The algorithm of computing two novel features based on Hilbert-Huang Transform is proposed in this section. The algorithm involves calculating the Instantaneous Frequency and Instantaneous Amplitude of the resulting coefficients after HHT [160], which are then used for feature extraction. An EEG-based recognition system using the proposed features was tested using two publicly available databases. Its accuracy and robustness was compared with a system using conventional wavelet-based features.

6.3.1 Motivation

Hilbert-Huang Transform was originally proposed by Huang *et al*. in 1998, since then this transform has received much attention in the signal processing field. Although HHT was particularly designed for analysing non-stationary signals, its theoretical justification is still debated: its effectiveness was however mostly demonstrated using various real-life experiments [161].

Standard HHT algorithm contains two steps: 1) the Empirical Mode Decomposition (EMD) is performed to obtain a series of Intrinsic Mode Functions (IMFs), 2) for each IMF the Hilbert Transform (HT) is performed to produce an Analytic Signal (AS). The detailed explanation of the EMD algorithm and HT for signal processing may be found in [62][160]. The first step of the HHT can be mathematically illustrated as follows: the input signal $x(t)$ is decomposed into a set of n IMFs, c_j, and a trend (residual) signal, r_n:

$$x(t) = \sum_{j=1}^{n} c_j + r_n \tag{6.2}$$

Each IMF is defined as to have symmetric upper and lower envelopes with the number of extremas and zero-crossings being equal or differing by one [62].

However, sometimes due to the "intermittency of the driving mechanisms", the EMD may trigger a phenomenon called "mode mixing" which causes IMFs to contain oscillations of drastically disparate scales [162]. Another drawback of EMD is that the solution of the decomposition is often not unique, i.e. the numbers of IMFs after the EMD can be vary for different windows (with the same window size) [62]. This may lead to a reduction of the algorithm's robustness in its application(s). To mitigate these shortcomings, Zhaohua *et al*. [163] proposed an EMD-based Noise-Assisted Data Analysis (NADA) method called Ensemble Empirical Mode Decomposition (EEMD). The features proposed for the EEG data analysis in this chapter is based on this improved EMD algorithm.

6.3.2 Ensemble Empirical Mode Decomposition

The EEMD may be described as follows: during the classic EMD process, a series of "finite, not infinitesimal, amplitude white noise" are added to the original signal [163]. This will create a uniformly distributed reference frame which may eventually exhaust all the other possible decomposing solutions (in standard EMD) and end up with a unique and stable ensemble mode number (number of the IMFs). According to Zhaohua *et al*. [163], the added white noise signals are chosen to be different in terms of amplitude. These uncorrelated white noise signals are likely to cancel each other out, along with the oscillations of drastically disparate scales which might cause the "mode mixing" phenomenon. The EEMD process proposed by [163] is as follows:

1. Add a white noise series to the input data;
2. Decompose the data along with the added white noise into IMFs;
3. Repeat step 1 and step 2 multiple times, but with different white noise series each time;
4. Obtain the (ensemble) means of corresponding IMFs of the decompositions as the final result.

The stoppage criterion of the EEMD (related to step 3) in this research is to set the number of the resulting modes to 12, based on some preliminary tests for the proposed EEG biometric system.

6.3.3 Instantaneous Frequency and Instantaneous Amplitude

For each IMF the HT is performed as described below. The Instantaneous Frequency (InsFreq) and Instantaneous Amplitude (InsAmp) functions can then be computed after performing the HHT algorithm. The idea of InsFreq and InsAmp are closely related to HT.

Denote the resulting IMF(s) after EMD by $x(t)$, the HT of $x(t)$ can be computed using (7.6):

$$y(t) = \frac{1}{\pi} P \int_{\tau}^{\infty} \frac{x(\tau)}{t-\tau} d\tau, \tag{6.3}$$

Where $y(t)$ is the result of HT and

$$P = \lim_{\varepsilon \to 0^{+}} \left[\int_{\tau}^{t-\varepsilon} \frac{x(\tau)}{t-\tau} d\tau + \int_{t-\varepsilon}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \right], \tag{6.4}$$

P is the Cauchy Principal Value defined by (6.4), due to the otherwise ill-defined function (6.3) while $\tau \to t$. Using HT to define the imaginary part of an Analytic Signal (AS), therefore, the AS can be defined according to the Euler's formula [164]:

$$z(t) = x(t) + iy(t) = A(t)e^{i\Theta(t)}, \tag{6.5}$$

In which

$$A(t) = \{x^2(t) + y^2(t)\}^{\frac{1}{2}} \text{ and } \Theta(t) = tan^{-1}\frac{y(t)}{x(t)}. \tag{6.6}$$

From this uniquely defined analytic pair, InsFreq and InsAmp can be computed. Though the definition of InsAmp, $A(t)$, is less controversial, the definition of InsFreq has been rather controversial [160]. For the proposed system, it is chosen to calculate the InsFreq, $\omega(t)$, as the derivative of Instantaneous Phase ($\Theta(t)$) [160]:

$$\omega(t) = \frac{d\Theta(t)}{dt}. \tag{6.7}$$

Intuitively the existence of InsFreq should be obvious, since for most of the non-stationary signals the power spectra density in frequency domain is changing over time, which makes a data-adaptive algorithm such as HHT a good option. This algorithm is used for feature extraction in the proposed system, as described in the following section.

6.3.4 System Design

To satisfy the need for maximum ease of deployment as a biometric system, only one electrode was employed for its performance evaluation. The EEG data was segmented using overlapped windows to capture informative events contained within the signal. The overall block diagram of the proposed system is illustrated in Fig. 6.8.

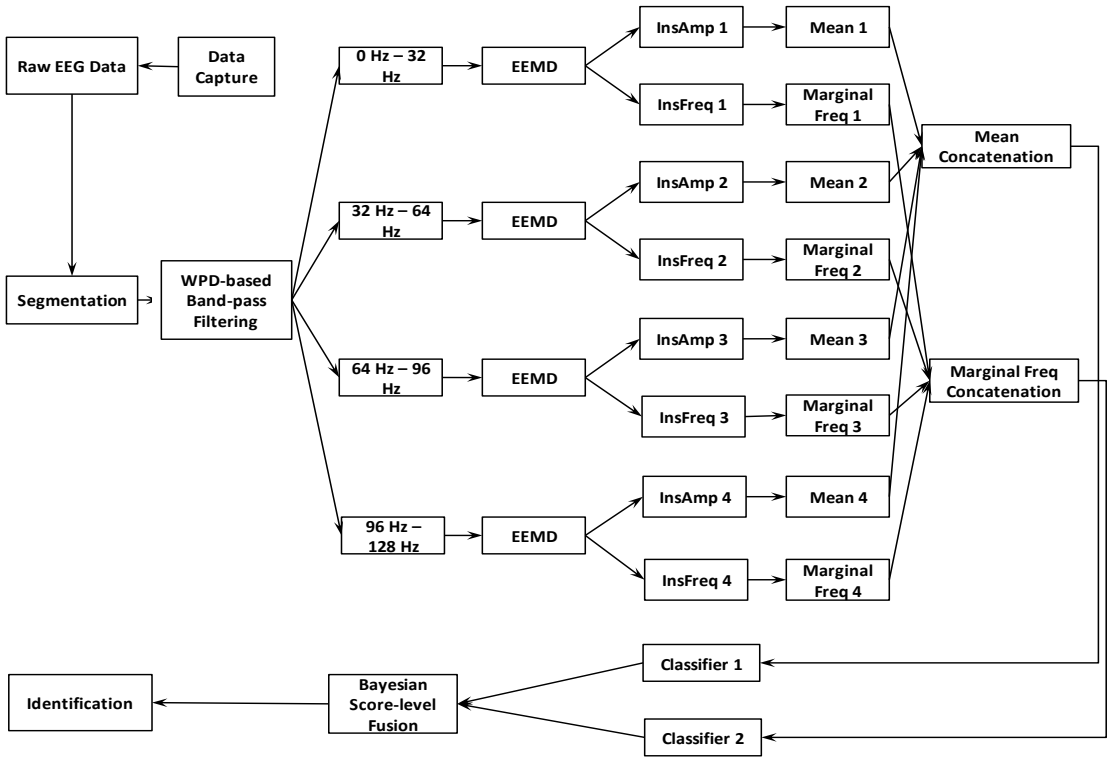


Fig. 6.8 System Block Diagram (frequency ranges according to VEP database)

A Wavelet Packet Decomposition algorithm [59] was then used to filter certain frequency bands (Fig. 6.8). The EEG signal was reconstructed into four equal sized

series in the time domain after WPD, each contained the frequency components from the selected bands: for the employed VEP database the four band ranges are 0~32 Hz, 32 Hz~64 Hz, 64 Hz~96 Hz and 96 Hz~128 Hz; for the MM/I dataset the corresponding bands range from 0~20 Hz, 20 Hz~40 Hz, 40 Hz~60 Hz and 60 Hz~80 Hz.

Each reconstructed band was processed using the EEMD algorithm to produce twelve empirical modes. The first and the last two modes were removed: the first mode simply was the replication of the original signal without decomposition, the last two modes have been found not to contain much of useful information.

The next step is to extract the InsFreq and InsAmp by performing the HT (using Eqn. (7.6)) on each mode. The function used in this work for computing the InsFreq was the derivative of the Instantaneous Phase, using Eqn. (6.7) in the previous section; the InsAmp $A(t)$ was calculated by Eqn. (6.6). Features were then generated by computing the marginal frequency of the InsFreq and the mean of InsAmp for each window. The choice of the mean and marginal frequency as the optimal feature generators was based on the testing of different operators including: Standard Deviation, Skewness, Kurtosis, Mode, median and Variance.

A feature-level concatenation after the computation of the InsFreq and the InsAmp was performed (Fig. 6.8): concatenating the data from four bands (each band contained nine dimensions – corresponding to the retained empirical modes) hence producing a 36 dimensions feature vector for the further processing. The same strategy was applied to InsAmp.

After this feature concatenation stage, the two sets of data from InsFreq and InsAmp were fed to two classifiers, respectively. Based on experiments using different classifiers, a Linear Discriminant Analysis-based classifier was used for InsAmp-based features, whereas the *k*-Nearest Neighbour (3-NN) was adopted for the InsFreq-based features.

The normalized scores generated by the two classifiers were then combined using a fusion rule based on Bayes' theorem, this fusion rule was used here to infer the joint

probability of InsAmp and InsFreq based classification results. The basic principle of Bayes' fusion rule can be defined as [165]:

$$P(F|A) = \frac{P(A|F)P(F)}{P(A)} = \frac{P(A|F)P(F)}{P(A|F)P(F)+P(A|(1-P(F)))P(1-P(F))}. \quad (6.8)$$

Here *P(F)* and *P(A)* are defined as the probability of correct classification of InsFreq-based and InsAmp-based features respectively. However, in this particular application, it is assumed that features from InsFreq and InsAmp are independent and identically distributed (i.i.d.), therefore the Eqn. (6.8) can be simplified as:

$$P(F|A) = \frac{P(A)P(F)}{P(A)P(F) + P(1-P(A))P(1-P(F))}. \quad (6.9)$$

Eqn. (6.9) is used to compute the fusion of the scores from InsFreq-based and InsAmp-based features.

6.3.5 Databases Description

The VEP Database [111] and the EEG Motor Movement/Imagery Dataset (MM/I Dataset) [15] are employed for the experiments.

In the evaluation to be reported in Section 7.3.6 only the middle run of Task 2 from MM/I database was employed for analysis. According to the preliminary research (tested with small number of samples/subjects), the data from Task 2 (imagine opening and closing left or right fist) was adopted for this experiment as it was found to provide the best biometric performance among the four tasks. The detailed instruction for each task may be found in [15].

Taking all the above-mentioned factors into consideration, due to the variation of the record lengths for each individual subject for the VEP database, only subjects with a recording length of at least 48 trials (48 seconds) were utilized, resulting in the inclusion of 118 out of 122 subjects. 105 out of 109 subjects were employed from Motor/Imagery database for the same reason; each subject with a recording length of at least 2 minutes.

Due to the nature of the tasks performed in different scenarios, for the VEP database, the Oz position (in the visual cortex) was adopted for analysis; whereas for the MMI

database the data from Cz (in the somatosensory cortex) position was used [14]. In the following section some evaluation results will be presented using these databases.

6.3.6 Experimental Results and Analysis

The first part of this section is devoted to analysing the impact of using different frequency bands on biometric identification performance. In the second part, the features adopted for the proposed system were evaluated and compared.

For the VEP database, 48 seconds of data is available and is sampled using windows of 16 seconds duration. The windows overlapped by two seconds. 12 such overlapped windows were concatenated to produce the data for frequency filtering. The same strategy was applied to the Motor Movement/Imagery Dataset: 120 seconds of data is available and is a sampled using window of 30 seconds' duration overlapped by two seconds. Due to the different sampling frequency of these two databases different window sizes have been used, in order to provide comparable window sizes (4096 sample/window for VEP database and 4800 samples/window for MM/I dataset) for the following feature extraction.

6.3.7 Frequency Band Sensitivity

The WPD was employed to separate the bands for analysis. The Daubechies 4 (db4) wavelet was adopted for the decomposition due to its relatively robust performance [108]. The performances achieved by using the reconstructed signals from each of the four bands were examined individually. 80% of the data were randomly selected for training and the rest for testing. The classifier used for InsAmp-based features was Linear Discriminate Analysis classifier (LDC) whereas for the InsFreq-based features the classifier used for most of the bands was the LDC except for 32 Hz~64 Hz band, which was classified by a 3-Nearest Neighbour (3-NN) classifier due to its relatively better performance.

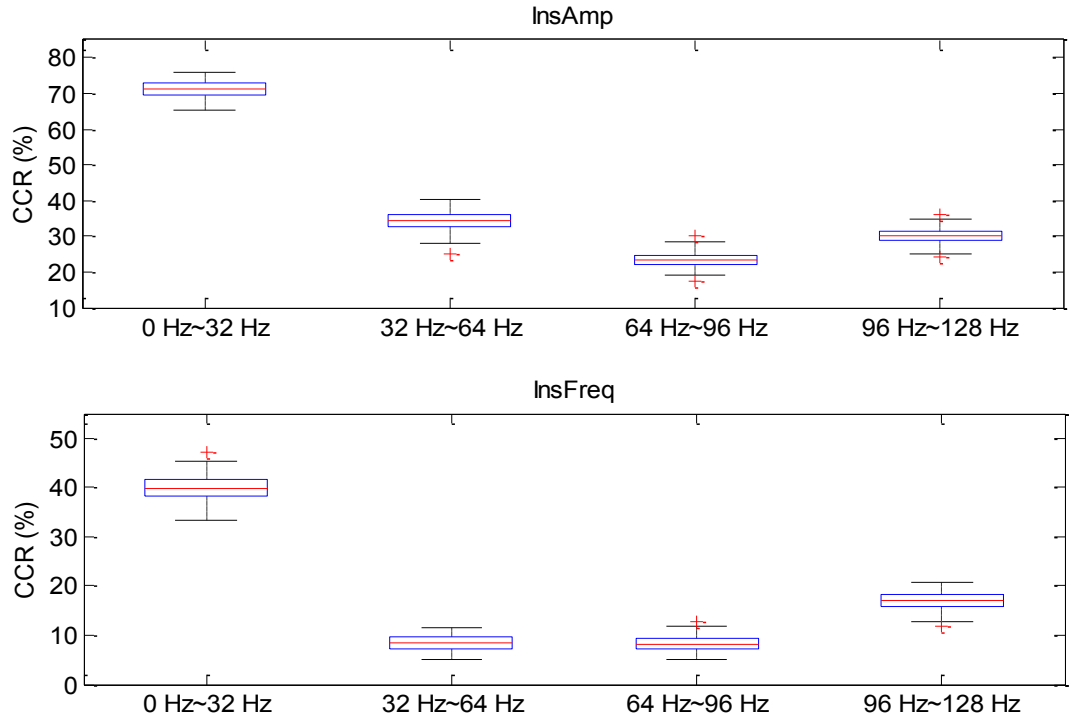


Fig. 6.9 InsAmp and InsFreq performances in bands, VEP database

Fig. 6.9 shows boxplots obtained from 100 random test attempts. All the boxplots in this work were generated from 100 attempts, each attempt with 20% of the data. The data from the frequency range between 0 to 32 Hz shows significantly better performance compared with the rest of the bands: This suggests that this low frequency band contained significantly higher biometric information compared with the other bands.

As the bands increase in frequency the identification performance is seen to degrade. The comparison between the InsAmp-based and InsFreq-based features clearly shows that InsAmp provides a much better performance in every corresponding band. It may be interesting to note that with InsFreq-based features, the performance is seen to gradually improve with increasing frequency. Due to the limitation of the cut-off frequency (128 Hz), it is impossible to analyse the performance at even higher frequency bands.

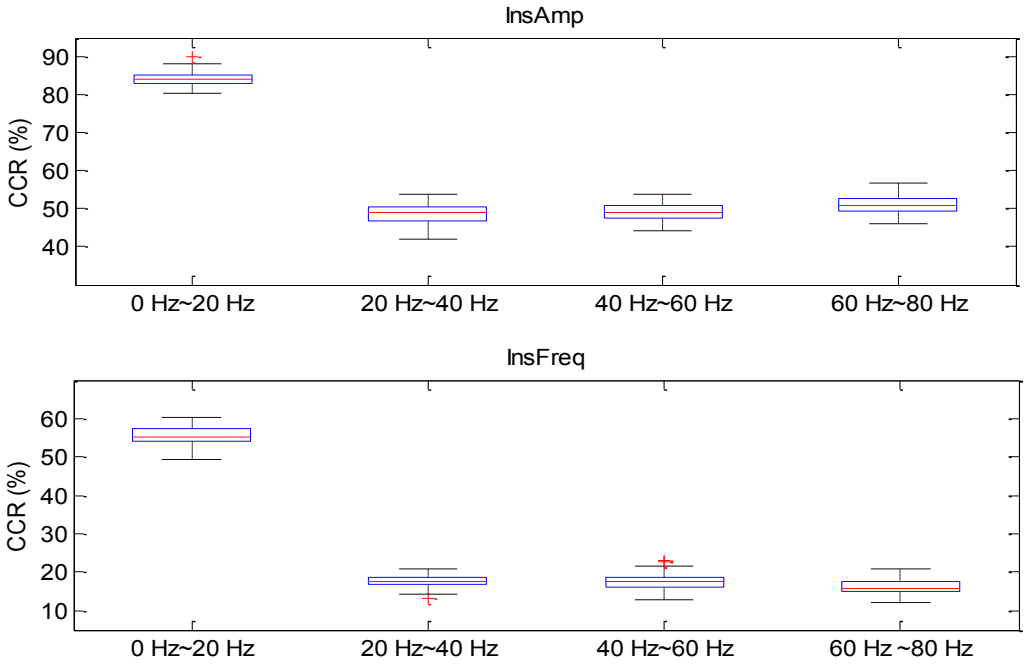


Fig. 6.10 InsAmp and InsFreq performances in bands, Motor Movement/Imagery database

The performances in terms of different bands for the Motor Movement/Imagery database are depicted in Fig. 6.10. This shows similar performance trends as that seen with the VEP database, only with higher Correct Recognition Rates (CRRs). The possible reasons for this improved performance could be: 1) the smaller number of subject for testing (105 subject versus 118 subjects); 2) the signal recording length is much longer (about 120 seconds versus 48 seconds); 3) perhaps motor imagery stimuli (Task 2) are better than visual stimuli at evoking distinctive electrical brain activity.

Although the higher frequency range provided much worse performance compared with the lower frequency bands, it is important to note that the high frequency bands may still contain useful complementary biometric information. The evidence for this can be found through the increasing of CRR when concatenated features from four bands are used (Fig. 6.11 and Fig. 6.12 in the following section).

6.3.8 Feature Sensitivity

It is clear that the lower frequency range yields much better biometric performance for both databases. Nevertheless, as it is shown in Fig. 6.11 while the information from

the four frequency bands is combined the overall identification performance further improved by about 20% compared with the single features from 0~32 Hz (Fig. 6.9). Significant improvement is also obtained in terms of InsFreq-based features after feature-level concatenation (Fig. 6.10).

After the concatenation of four bands from InsAmp and InsFreq based features respectively, the concatenated features were fed to two classifiers for training, individually (the two classifiers are not necessarily the same). The InsAmp and InsFreq score-fused performance further improved about 5% compared with the performance achieved using only the InsAmp feature (Fig. 6.11). The same fusion scheme also improved the overall performance for the Motor Movement/Imagery database (Fig. 6.12).

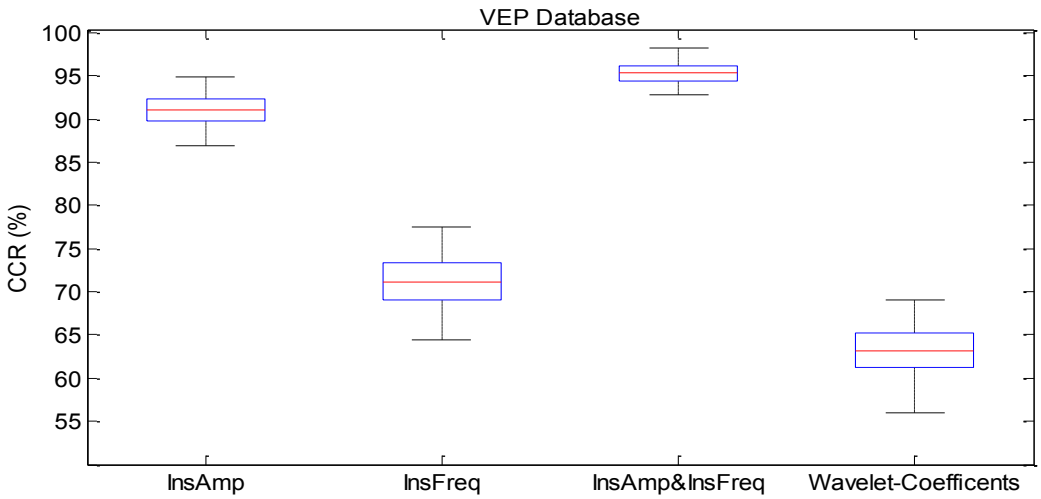


Fig. 6.11 The comparison of four feature extraction methods, VEP database

The right-most boxplot of the Fig. 6.11 and Fig. 6.12 depicted the performances of wavelet-based method described in [108] for each of the database. Certain parameters (number of observations and dimensions, for example) were set to be comparable with the HHT-based method for the sake of comparison. One can clearly see that the wavelet-based method provided comparable performance with the InsFreq-based algorithm, but when compared to the proposed InsAmp-InsFreq system it does not perform quite well (Fig. 6.11 and Fig. 6.12). It should be noticed that, though it seems the performance from the InsAmp-based feature alone is not much worse than the

InsAmp-InsFreq combined features, yet the variance of CRR using the fusion method (Eqn. (6.9)) is clearly smaller. This may indicate that as the number of the subjects is further increased, the advantage of the InsAmp-InsFreq system will be even more significant.

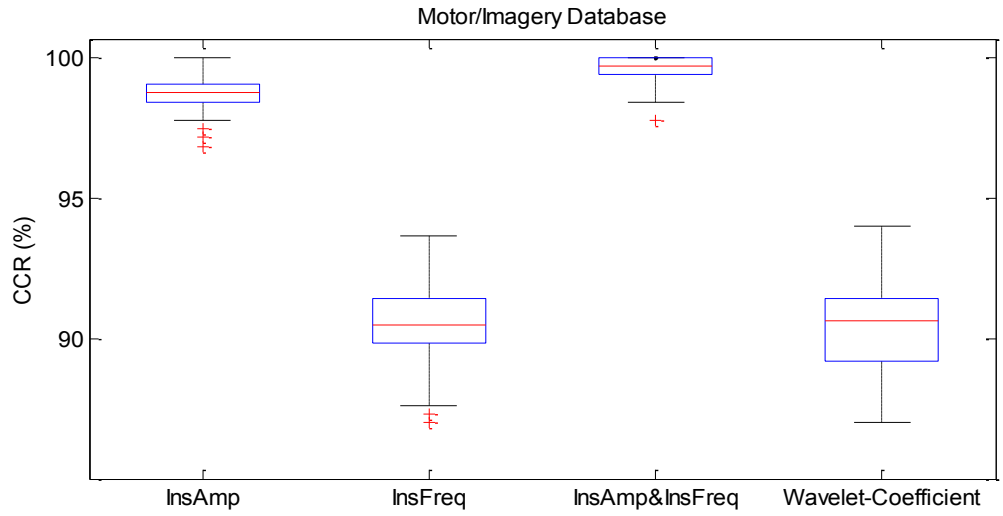


Fig. 6.12 The comparison of four feature extraction methods, Motor Movement/Imagery database

6.3.9 Comparative Overall Performances

The performances of some previously published approaches by other authors were compared with the proposed system in this section. The mean performances of the proposed system tested by different features are depicted in Table 6.4. It is clear that InsAmp-based features perform better than InsFreq-based features. The fusion of the two further improves the identification rate. It seems that the InsFreq-based features and the previously proposed wavelet-based features not only performed worse but are also less robust: the performance degraded a lot while the database changed (consider the second and fourth columns of the Table 6.4).

Table 6.4 Comparison of different features

Single Electrode	InsAmp	InsFreq	InsAmp & InsFreq	Wavelet Method
VEP Database (Oz)	91.02%	70.56%	95.88%	63.19%
MM/I Dataset (Cz)	98.63%	90.59%	99.50%	90.43%

The comparison of difference reported-systems in terms of the identification rate using the VEP database is illustrated in Table 7.5. The most prominent advantage of the proposed system is the number of electrode: a single electrode system with more than 95% of CRR for 118 subjects by far surpasses the performance reported for previous systems.

Table 6.5 Comparison of different systems

VEP Database	Proposed System	Yazdani *et al*. [74]	Brigham *et al*. [56]	Huang *et al*. [66]
Number of Electrodes	1	64	64	64
Number of Subjects	118	20	120	116
CRR	95.88%	100%	98.96%	95.1%

The VEP database with its considerably shorter recording lengths and more subjects provides a challenging test for an EEG-based identification system. Therefore, the result of the proposed system using only a single electrode for the VEP database, reaching a CRR of about 96% is particularly interesting.

Though these results are encouraging, the challenge still remains to establish the usability and robustness of EEG-based biometrics, especially using single low-cost sensors and short duration training and testing regimes. One of the biggest limitations of the current databases is the fact that there is no significant time separation between training and testing sessions. The VEP database contains only one session, and the MM/I database, though containing three runs, has these separated by an interval of only a couple of minutes. Ideally multiple sessions with time intervals of several days would be required to establish the stability of EEG signals as a biometric modality. The collection of such data to address this challenge is an important focus of activity for research in this field, its related exploration is also a major task included in this thesis (see Chapter 3 and Chapter 8).

6.4 Feature Comparison and Discussion

In this section the biometric recognition performances of the proposed features are investigated to compare their effectiveness using the publicly available MM/I dataset.

The results are the average performance of three fold non-overlap cross-validation: each fold correspond to a 2 minutes of EEG recording. Only the data of the electrode in Oz location was employed in these tests. It should be mentioned that the recognition rates reported here are for comparison purposes; better biometric performance by employing a novel feature classification algorithm under a similar experimental scheme is reported in Chapter 8.

All of the four newly proposed features are tested in an identification scenario. Another five conventional wavelet-based features are tested as well for the sake of objective comparison. Each feature is computed using 4800 samples (30 seconds) per window; a simple nearest neighbour (1-NN) classifier is employed to compare the effectiveness of the features. The Cumulative Match Characteristic (CMC) curves generated using nine features are illustrated in Fig. 6.13.

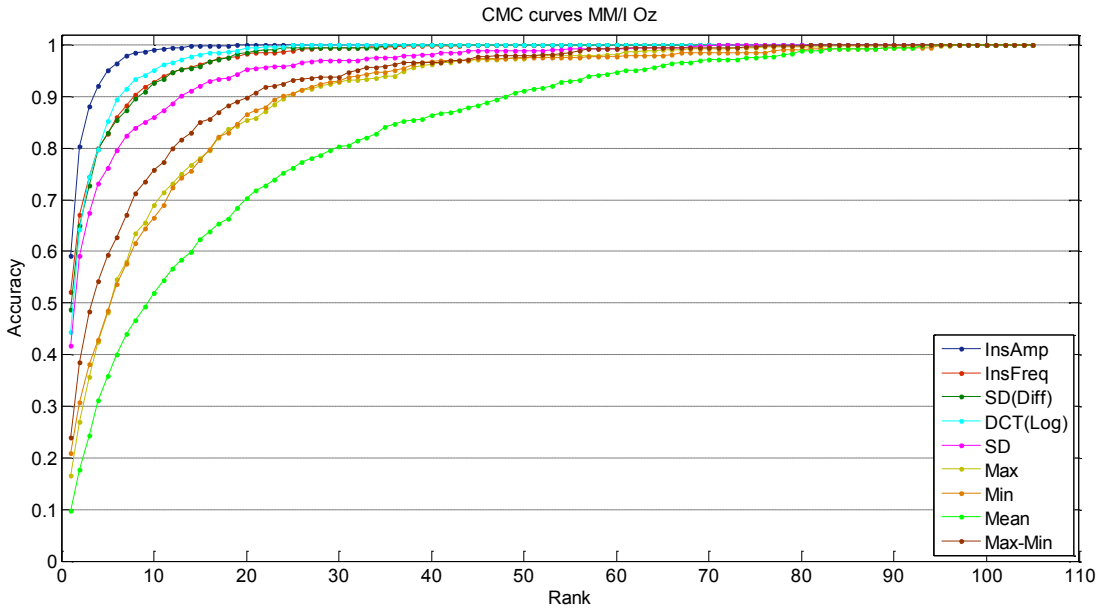


Fig. 6.13 Recognition comparison of multiple features

The InsAmp and InsFreq features introduced in Section 6.3 provided better rank one performance than the two new wavelet-based features presented in Section 6.1 and Section 6.2. The CMC curve of InsAmp feature provided both the highest accuracy. The InsFreq feature and the *SD*(*diff*) feature provided comparable learning speeds (Fig. 6.13), despite of the slightly higher rank one accuracy from InsFreq feature. The wavelet-DCT feature, though delivered only the lowest rank one identification rate

amongst the four newly proposed features, the rather fast learning speed can be observed compared with InsFreq and *SD*(*diff*) features.

The five conventional wavelet-based features are based on the resulting wavelet coefficients: the purpose is to reduce the dimension of the feature vector in the meantime hopefully still capture useful biometric information. Interestingly, the Maximum and Minimum strategies shared similar performances, whereas the mean of the wavelet coefficients seems to have ignored the possible temporal information of EEG signals due to its obviously worst performance of all (Fig. 6.13). Considering this notion, attempt was made to capture the temporal characteristics of EEG signals by further deriving a simple feature which combines the Max and Min strategies: Max-Min feature (the maximum coefficient minus minimum coefficient per window). The related curve in Fig. 6.13 indeed shows an improvement: the Max-Min feature better performed the Max and Min strategies in terms of both the rank one identification and learning speed.

6.5 Conclusion

Four novel features for EEG biometric recognition were proposed and their performances have been experimentally evaluated in this chapter. Two HHT-based features and two Wavelet-based features were investigated and compared using two publicly available databases. The HHT-based features provided better overall performance than the wavelet-based features, however, the computation of HHT features have been found to be much more time consuming than wavelet features (more than 60 times slower) due to the looping process of the EEMD algorithm.

In the following chapter (Chapter 7), a novel quality filtering algorithm is proposed. The recognition accuracy is improved by removing the less informative signals segments.

Chapter 7

Quality Filtering Algorithm

This chapter introduces a biometric person recognition system based on EEG signals incorporating a novel quality filtering strategy to find and utilize the most informative segments (windows) using the concept of Sample Entropy (SampEn) [166]. A sliding-window segmentation scheme and Wavelet Packet Decomposition (WPD) were employed for the primary feature extraction before the quality filtering. After the quality filtering, the preserved segment windows were then used to extract the secondary features that were in turn classified using a Linear Discriminant Classifier (LDC).

The method was tested using the data of the whole publicly available MM/I dataset. An average identification accuracy rate of more than 90% was achieved for 109 subjects using eight selected electrodes, utilizing the highest quality training data obtained from 4 minutes of recording for each subject. The experimental results indicate a 5% of accuracy improvement by applying the SampEn filtering method. The result also shows the SampEn is sensitive to identification performance: using the low SampEn-corresponded features for the same classifier training provided low performance.

This chapter is organised as follows: the general EEG data acquisition scheme will be presented in Section 7.1, along with the electrodes positioning and the block diagram of the biometric recognition system. The rationale of the feature extraction algorithm used in the proposed system will be introduced in Section 7.2. Section 7.3 outlines the novel entropy-based method for data quality filtering, followed by the experimental results as well as the tests for optimizing the system parameters in Section 7.4. Section 7.5 provides the summary and suggestion for further work.

7.1 Experimental Scheme

The proposed biometric data acquisition system is based on measuring the evoked response of users while they are confronted with a stimulus and asked to perform a mental task. The overall setting was introduced in Chapter 3. The "MM/I database was used for evaluating the proposed algorithm and the performance of the system in an identification (one-to-many recognition) scenario.

The proposed system is trained and evaluated using the data obtained from Task 4. The reason for adopting Task 4 (motor imagery task for both hands and feet) is that the motor imagery task might better avoid the possible contamination by other bioelectrical signals such as Electromyography (EMG) noise. Due to the need for the ease of deployment as a biometric modality, EEG signals from up to eight electrodes were used and their positions are clustered around the centre of the motor cortex: FC_1, FC_2, C_3, Cz, C_4, CP_3, CPz, and CP_4 [16]. The electrode positions for this experiment followed the Scheme III introduced in Chapter 4.

The block diagram of the proposed system is shown in Fig. 7.1. The Wavelet Packet Decomposition [59] is used to generate the primary features (windows with wavelet coefficients). An entropy-based measurement method is designed to select optimal windows for generating secondary features which are then passed to a classifier. *M* (number of windows), *K* (kept number of windows) and *L* (the starting window for preservation after window sorting by quality) are three parameters for controlling the system performance and to be described in the following sections.

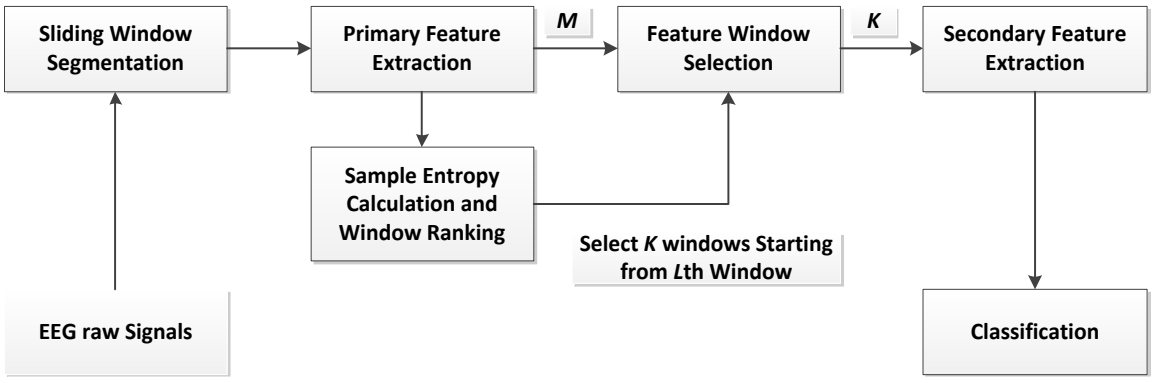


Fig. 7.1 Block diagram of the system

7.2 Feature Extraction

One goal of pre-processing and feature extraction, especially when EEG signals are used for biometric recognition, is to remove the unwanted signal segments (such as signals generated by eye blinks and heartbeats) and only preserve the parts with essential identity-bearing information for classification. For the proposed system, Wavelet Packet Decomposition is utilized for the first stage of feature generation [136]. This is followed by a novel method for quality measurement is designed to select and use only a small amount of data segments which are most likely to provide a correct recognition before feature extraction.

Each run (individual recording) of the EEG data in MM/I dataset lasted approximately 2 minutes, which are divided into several windows with each contained 960 samples (6 seconds) in this experiment. WPD is performed on every window, from each window 9 bands of wavelet coefficients are generated which correspond to typical EEG bands in the wavelet domain. The two highest frequency bands that are produced by wavelet decomposition are discarded and not used for further analysis, as they correspond to high frequency (close to the cut-off frequency) signal components and likely been contaminated by noise.

7.3 Sample Entropy for Quality Measurement

This section explains the novel entropy-based filtering method in details. The motivation of adopting Sample Entropy for quality measurement is presented first, followed by the mathematical definition of Sample Entropy, and finally the scheme of using it for EEG filtering in the proposed system is described.

7.3.1 Motivation for using Entropy as a Measure of Quality

As a stochastic time variant series, the power spectra of typical EEG signal vary over time depending on what brain functions are being performed. Also similar to the speaker recognition, it is unlikely the whole signal series is equally informative for the purpose of identity recognition. In order to reduce the amount of the data used for processing and improve its quality, it is necessary to find a strategy to extract the

most useful segments of the data and discard the relatively less-informative portions of it. EEG signal is often contaminated by other bio-signals during the data collecting process, such as the electrical signals generated by the activity of the heart and muscular movements [167].

Entropy of the EEG signal has been used as a feature to identify the seizures in epileptic patients. It has been reported that during seizure, patients' brains generate lower entropy EEG signals than for healthy people. This implies that the healthy brain signal may possess less regularity than a brain during seizure [11]. In that reported experiment, three different types of entropies (Shannon entropy, Sample entropy and Log entropy) were calculated for EEG segments and all of them showed such a trend. Liang *et al*. [168] later reported that the entropy of the EEG could also be used to identify the sleeping stages. They measured the EEG signal on an epoch-by-epoch basis, using multi-scale entropy analysis (MSE) and noticed that the "entropy values monotonically decrease from awake to deep sleep" [168]. These results suggest that entropy may be used to measure the level of brain activity from EEG signals in healthy human brain functions.

The more active the brain is with cognitive/motor functions, the more unpredictable the EEG signal is likely to be, hence the higher the corresponding entropy value. Based on this hypothesis, in this work the Sample Entropy as a measure is proposed for EEG signal quality filtering in biometric recognition.

7.3.2 Definition of Sample Entropy

Sample entropy (SampEn) examines a time series and assigns a non-negative number to the sequence, number with larger value corresponding to greater irregularity being present in the data [166]. Two input parameters, a run length m and a tolerance threshold r must be specified for its calculation [169]. For a time series of length N, the function SampEn (m, r, N) compares sequences of length m and $m + 1$ in the time series and measures the similarity. A lower value of SampEn indicates more self-similarity in the time series [170]. Formally, given a time series with N data points

as $\{x(n)\} = x(1), x(2), \ldots, x(N)$, SampEn may be computed by the following steps [166]:

1. Form multiple vector sequences of size m, $X_m(1), \ldots, X_m(N-m+1)$, defined by $X_m(i) = \{x_m(i), x_m(i+1), \ldots, x_m(i+m-1)\}$, where $1 \leq i \leq N-m+1$. These vectors represent m consecutive x values, starting with the i^{th} point.
2. Define the distance between vectors $X_m(i)$ and $X_m(j)$, $d[X_m(i), X_m(j)]$, as the absolute maximum difference between their scalar components:

$$d[X_m(i), X_m(j)] = max_{k=0,\ldots,m-1}(|x(i+k) - x(j+k)|) \tag{7.1}$$

For a given $X_m(i)$, count the number of j ($1 \leq j \leq N-m, j \neq i$), denoted as B_i, such that the distance between $X_m(i)$ and $X_m(j)$ is less than or equal to r. Then, for $1 \leq i \leq N-m$:

$$B_i^m(r) = \frac{1}{N-m-1} B_i \tag{7.2}$$

Therefore,

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \tag{7.3}$$

3. Increase the dimension to $m+1$ and denote A_i such that the distance between $X_{m+1}(i)$ and $X_{m+1}(j)$ is less than or equal to r, for j ($1 \leq j \leq N-m, j \neq i$). Then, $A_i^m(r)$ is defined as

$$A_i^m(r) = \frac{1}{N-m-1} A_i \tag{7.4}$$

Set $A^m(r)$ as

$$A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \tag{7.5}$$

Thus, $B^m(r)$ is the probability that two sequences will match for m points, whereas $A^m(r)$ is the probability that two sequences will match for $m+1$ points.

4. SampEn is then calculates by:

$$SampEn(m, r, N) = -ln\left[\frac{A^m(r)}{B^m(r)}\right] \quad (7.6)$$

7.3.3 Using Sample Entropy to Filter EEG Data

After calculating the WPD for each window of the time series, the sample entropy of each wavelet window (each with multiple coefficients) is computed. In this experiment, each recording run of approximately 2 minutes is segmented into windows of 960 samples (6 seconds duration) using a sliding window approach (overlapping), thus producing 760 windows for every EEG band per electrode. More generally, the number of windows generated per band per electrode can be defined as a system parameter, *M*, as is shown in Fig. 7.2. These coefficient windows are then fed to the SampEn calculation module which ranks the windows by ordering their entropy values from the highest entropy window to the lowest. For each band, preserve *K* out of *M* entropy values in order to shrink the data scale and remove the information-poor windows. In the experiments that follow, only about 1/10 of the data (80 out of 760 windows) is used for the secondary feature extraction. The standard deviation (SD) of the wavelet coefficients from the selected windows are then calculated to serve as the features for classification.

It has been suggested that the choice of *r* is important in the calculation of SampEn: if it is set at too high a value, detailed system information may be lost and if it is set at too small a value, poor conditional probability estimates might result [166]. These predictions may affect the selected (filtering) candidate windows of the proposed algorithm. As a rule of thumb, the tolerance threshold *r* is set to 1 and run length *m* is set to 2 [166] for the experiments to be reported in Section 7.4.

7.4 Experimental Results and Evaluations

The illustration of how the system handles different EEG frequency bands was shown in Fig. 7.2. During the training phase, for every electrode per subject $M = 760$ windows were fed for the WPD transform. Each window was decomposed into nine bands of wavelet coefficients. Since eight electrodes were used in the experiment,

before the entropy filtering stage a total of 8×760×9 = 54720 wavelet coefficient windows were generated for each subject. Next, the Sample Entropy is calculated for each window and the windows are sorted in descending order according their corresponding SampEn values.

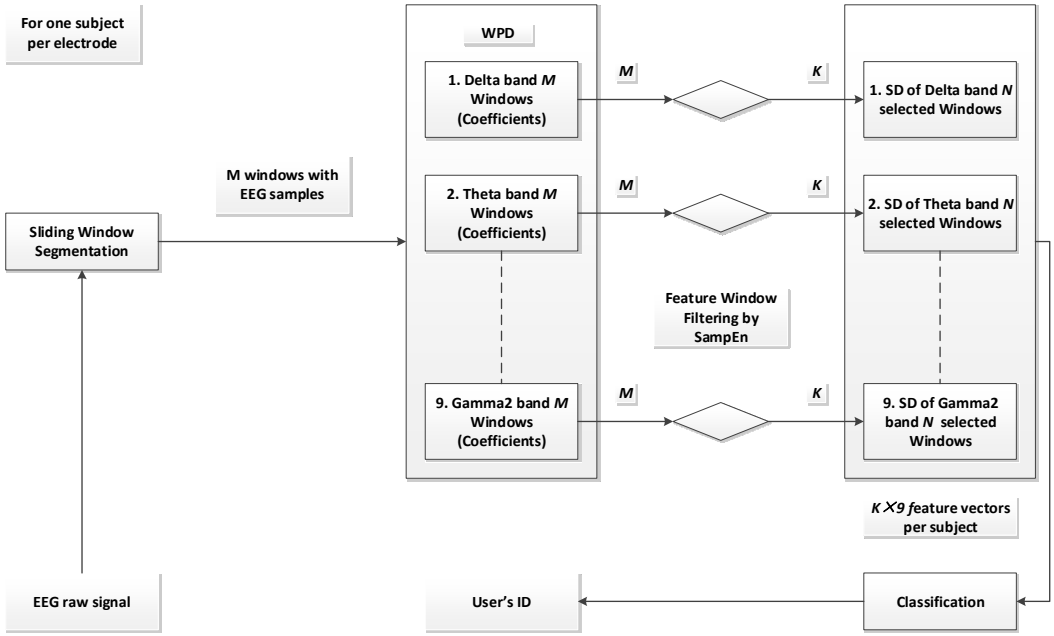


Fig. 7.2 Diagram of system algorithm

Out of M (=760) windows from each electrode per band, K (=80) windows (hereby referred to as "observations") are retained as a contiguous range from the SampEn-sorted list of windows to preserve the most information-bearing part of the data. This amounted to roughly 10% of the whole data. After this quality measurement and screening stage, for every subject only 5760 out of 54720 windows were kept and the standard deviation (σ) of each coefficient window is calculated and used as features for classification, using a normalised linear discriminant classifier (LDC, [171]). This choice of classifier was based on tests and comparisons with several other classifiers using the MM/I dataset (Support Vector Machines with different kernels, k-Nearest Neighbour classifiers, kernel- LDC and kernel-k-NN [171]).

7.4.1 Entropy Filtering Optimization

Different contiguous ranges of windows from the entropy-sorted list of coefficient windows were extracted and used to filter the training data, only those windows

within the selected range were used for the classifier training. As is shown in Fig. 7.4, the first range testing is for the windows with the highest SampEn values (rank 1 to 80). The SampEn values of their corresponding windows (760 windows in all) are sorted in monotonically descend order. The system was trained with only these windows (80 windows per point, the selected windows indicated by red point in the figure) retained and when tested with data from Run 2 the identification accuracy results are about 90% for 109 subjects.

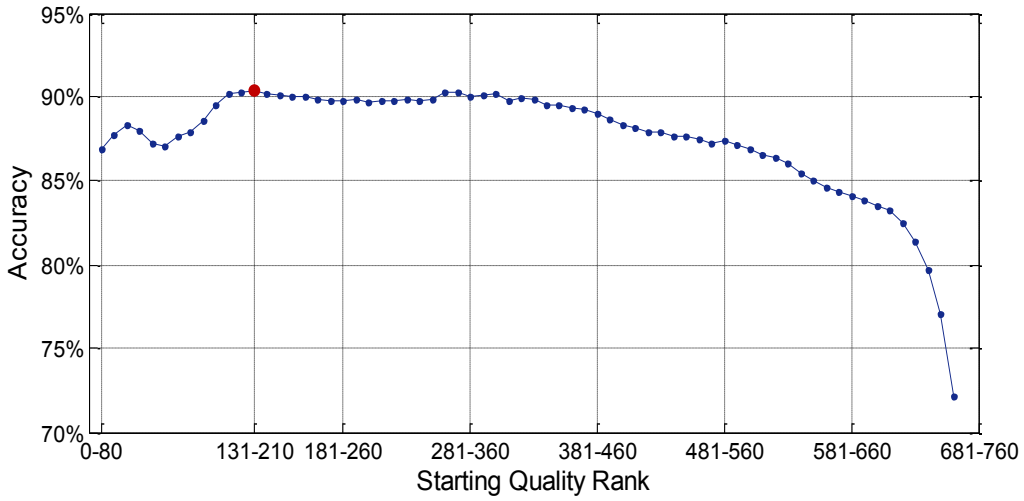


Fig. 7.3 Identification accuracy as a function of the quality rank range starting value (L) selected for inclusion in training. K=80

The range of selected ranks was then moved every 10 ranks but always containing 80 ranks within the range. Fig. 7.3 shows the resulting accuracy for different starting ranges, the best performance was achieved while L = [131, 210]. These results may not fully fit the original conjecture, as the best identification performance does not precisely correspond to the segments with the highest SampEn. However, there is clearly a range of high-entropy ranks (131th to 360th range) which still provide high biometric performance. It is possible that the highest ranking windows correspond to activities that do not carry identity information. As each window lasts 6 seconds, it may include 5 to 6 cycles of motor actions (e.g. opening and closing of hands). This could be considered as a relatively regular function with moderate SampEn values. Testing results of different parameter K suggests that the features corresponding to the highest around 15% ranking Sample Entropy values should be discarded.

7.4.2 Performance as a Function of Test Segment Duration

Fig. 7.4 depicts the degradation of the accuracy rate for identification when the testing duration t is reduced. These tests are all based on observation K=10 and starting value L=131 (red point in Fig. 7.3), the average results utilizing Run 2 (or part of it) for testing and the other two Runs for training. Hence, 4 minutes of recording is used for training the classifier. Dropping the test segmenting duration by a factor of four from 120 seconds to 30 seconds results in a loss of accuracy of only 3.48%; dropping the test duration all the way to just 6 seconds, a 20 times reduction, results in a drop in accuracy of less than 9%. It should be clarified that with the dropping of the test duration, the window sizes also dropping accordingly, which could be observed from the numbers of the dots in Fig. 7.4.

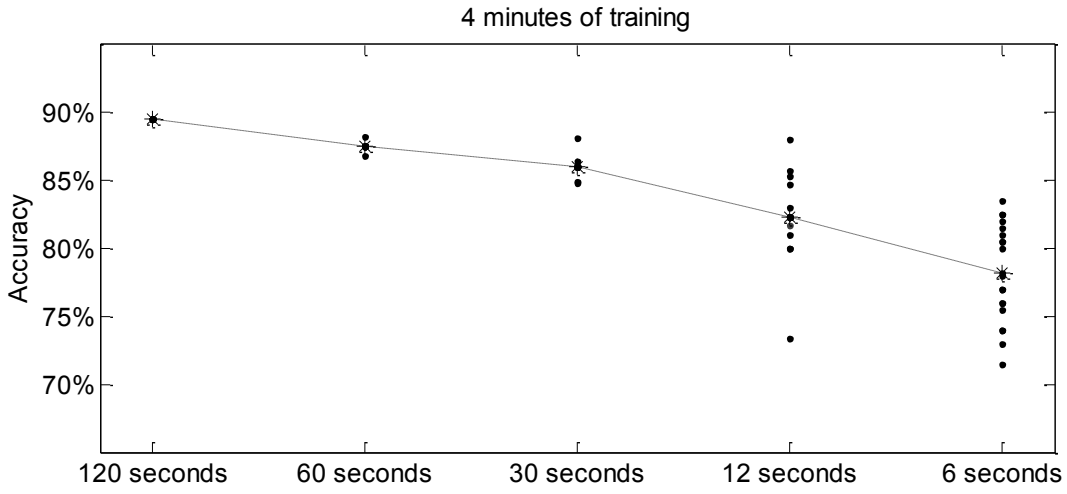


Fig. 7.4 Impact of the testing duration on average identification accuracy with K=10 and L=131

7.4.3 Impact of Entropy Filtering

To illustrate the impact of the proposed entropy filtering method, in this section a number of schemes with different parameter setting are compared. Five schemes are tested (Table 7.1), the scheme I to III are employed entropy filtering algorithm, scheme IV and V are methods only utilizing wavelet decomposition.

Table 7.1 Comparison of different schemes with and without entropy filtering

Schemes	L: Rank starting value	K: Preserved observations	Accuracy
I. Highest entropy	**1**	**80**	**87.0%**
II. Highest performance	**131**	**80**	**90.4%**
III. Lowest entropy	**681**	**80**	**74.8%**
IV. No entropy filtering	**1**	**760**	**86.5%**
V. No entropy filtering	**1**	**80**	**85.6%**

The results presented in Table 7.1 indicate that, given the chosen system parameters, the entropy filtering method improves the recognition performance by around 5% compared to using no entropy filtering at all. The results for Scheme III suggest that the low entropy-windows contain significantly less biometric information, which indicate the link between the sample entropy of the signal and its biometric potential.

7.5 Summary

The study in this chapter explored the concept of quality for EEG signals used for biometric person identification. A novel system was presented where a measure of signal quality, the Sample Entropy, was used to evaluate and filter the data available for biometric recognition. For a 109 subject database an identification rate of more than 90% was achieved. The results indicate comparative performance with other published methods while promising the possibility of being able to handle large number of subjects using data from fewer electrodes. Further work will focus on optimizing the system parameters separately for different frequency bands and increasing the amount of data used for system evaluation.

In the next chapter (Chapter 8), a novel instance-based template reconstruction learning algorithm is to be proposed, which achieved considerable recognition improvement.

Chapter 8

Instance-based Template Reconstruction Algorithm

In the machine learning family, the instance-based learning algorithms are designed to compare the new instances with the instances statically stored in the memory (lazy learning) [165]. Typical instance-based learning algorithms include *k*-nearest neighbour (*k*-NN) and support vector machine (SVM) [172]. One advantage of the instance-based methods over other learning methods (such as eager learning algorithms) is the ability of adaptably modelling the training set: with the increasing of the training data, the prototypes of interest could be updated, hence potentially maintain the performance. However, the increasing of the prototypes also introduces the complexity of the training set: to deal with this shortcoming of these algorithms, a series of instance reduction (prototype selection) algorithms are developed [173]. Many prototype selection algorithms have been developed with a range of objectives in mind, including 1) storage reduction, 2) speed increase, 3) improved generalization accuracy, 4) noise tolerance, 5) increase learning speed and 6) incremental [174].

One of the earliest instance reduction algorithms follows the so-called Condensed Nearest Neighbour (CNN) selecting rule, proposed by Hart [175]. The core notion of this algorithm is to incrementally test the instances of the training set *T* using a subset of *T*. By preserving only the correctly classified instances in *T*, a new training subset in which the instances are better discriminated is used for classification. However, this method could not identify the possible noise instances due to the equal-weighted comparison process during the classification [172]. Much work has been done since then: summaries of relevant studies are listed in a number of review papers, such as [172] [174] [176].

A novel instance-based template reconstruction learning algorithm is proposed in this work. The effectiveness of the algorithm is tested using two EEG databases to demonstrate its robustness.

The structure of this chapter is organized as follows: Section 8.1 is devoted to the motivation for developing the algorithm, which includes a list of influential factors that should be considered while designing an EEG biometric system. The detailed description of the algorithm is given in Section 8.2 and its subsections. In Section 8.3 an illustrative step-by-step walk-through of the algorithm is presented using real EEG data. Comparison between the proposed method and two other popular pattern recognition algorithms is also provided, namely the Support Vector Machine (SVM) and the *k*-Nearest Neighbour (*k*-NN) algorithms. In Section 8.4, the performance of the proposed algorithm is evaluated for a number of pattern recognition tasks. Section 8.5 provides a discussion of the results and overall conclusions.

8.1 Motivation

The proposed algorithm is aimed to improve the pattern recognition performance with EEG signals using very limited data, in terms of quality, duration and quantity compared to the number of different classes that need to be distinguished. Several issues need to be considered and improved simultaneously during experiments in order to achieve its successful real-world biometric application using EEG signals. Besides the recognition performance, a couple of important factors in evaluating an EEG-based biometric system that affect their usability are listed as follows:

1) Number of electrode(s) involved for processing
2) Recording lengths needed for training and query sets
3) Impact of template ageing on performance
4) The level of difficulty in setting up the sensor system.

Most of the work reported in the Literature Review chapter (Chapter 2) could not fully address all of these requirements. In this chapter, the above-listed factors are considered with the aim of moving towards implementations of EEG biometric system outside the controlled laboratory environments: an easily-deployed, commercially available low-cost single dry sensor system [17] is employed for the real-time data collection and classification; a novel instance-based learning algorithm is developed to alleviate template ageing effects and allow the reduction of the overall

data capturing time, which was tested using MM/I dataset and the Mobile Sensor Database.

The proposed algorithm in this chapter seeks to 1) reduce the impact of the template ageing effect on performance of EEG signals and, 2) improve the recognition performance using very limited data, as opposed to the training data accumulation scheme proposed in Chapter 5, which achieved performance improvement by increasing the training set volume.

The main task for pattern recognition is to determine the membership (class) of the instances/patterns. The level of between-class (inter-class) similarity and within-class (intra-class) similarity are two critical factors in evaluating the instance/pattern distribution [144]. For pattern recognition in general, high inter-class similarity may lead to high FAR and low intra-class similarity may leads to high FRR. To achieve good recognition performance, instances with low inter-class similarity characteristics and high intra-class similarity characteristics are usually preferable.

Besides the inter-class discrimination (for most of the physical biometric modalities), the intra-class variance (reflecting by template ageing effects) is quite an influential factor for EEG-based pattern recognition: even a couple of minutes' time interval between independent recordings may lead considerable template ageing effects. Therefore, instance selection becomes a necessity during pattern modelling in order to achieve both good between and within class discriminations. To clearly demonstrate the effectiveness of the proposed algorithm, EEG signals provide an appropriate test case.

Furthermore, some preliminary investigations (in Chapter 4) revealed that the patterns of EEG signals (from different classes) employed in this work overlap considerably and are sparsely distributed in feature space. Therefore, filtering out the less informative instances/patterns is likely to have a substantial effect in enhancing the recognition performance.

The novel template reconstruction/recreation learning algorithm which is presented below contains two stages: *Stage 1* and *Stage 2* of the algorithm are designed to cope with the previously mentioned inter-class and intra-class problems, respectively.

8.2 Instance-based Template Reconstruction Learning Algorithm (ITR)

The proposed algorithm is presented in Section 8.2.1. To facilitate its understanding, an illustrative example of the proposed algorithm is provided in Section 8.2.2. Finally in Section 8.2.3 several remarks are presented for the better understanding of the algorithm's properties.

8.2.1 Algorithm

For an *N*-class identification problem, denote the training set of instances/patterns as T and the query/test set as Q. Assuming the same number of instances is available for each of the *N* classes (for the simplicity of the algorithm illustration), denote the number of instances as *I* per class, with each instance being composed of an *L*-dimensional feature vector. The data elements in the training set may, therefore, be represented by a three-dimensional matrix $T_{i,n,l}$, where i denotes the index of each element of an instance (measurement/pattern), n denotes the class to which that instance belongs and l denotes the feature dimension for that instance. The complete training feature set is denoted as $T_{:,:,:}$. The algorithm operates on each feature dimension, l, separately, i.e. $T_{I,N,l}(1 \leq l \leq L)$ as indicated in Fig. 8.1 (the query set follows the same notation). In the following algorithm description, the symbols with lower case subscripts, such as $T_{i,n,l}$ and $S_{i,n,l}$, are used to indicate the location of the element in the matrix; the symbol with capitalized subscripts, such as $T_{I,N,L}$ and $S_{I,N,L}$ are used to indicate the entire matrix/sub-matrix. The distance(s) d calculated as part of the algorithm is the Euclidean distance.

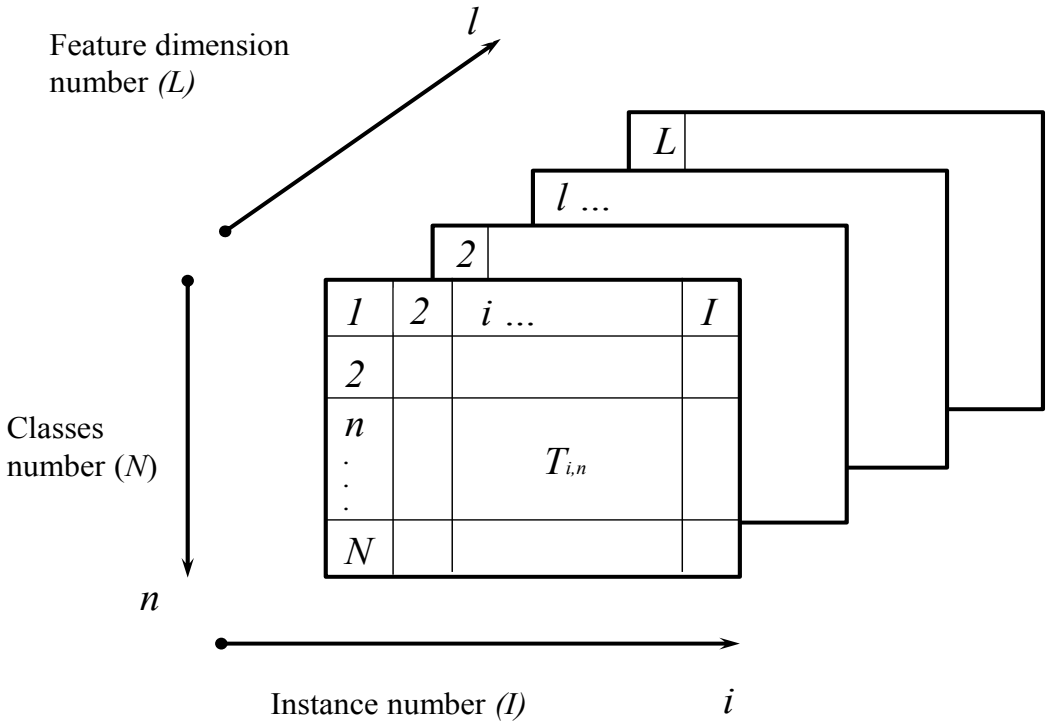


Fig. 8.1 Training set matrix $\boldsymbol{T_{I,N,L}}$, where $\boldsymbol{1 \le i \le I, 1 \le n \le N, 1 \le l \le L}$.

The overall block diagram of the proposed algorithm is illustrated by Fig. 8.2. The circled numbers indicates the processing order of the algorithm. The circled numbers indicate the order of the algorithm modules; the purpose of this diagram is to illustrate the three circles involved in the two stages.

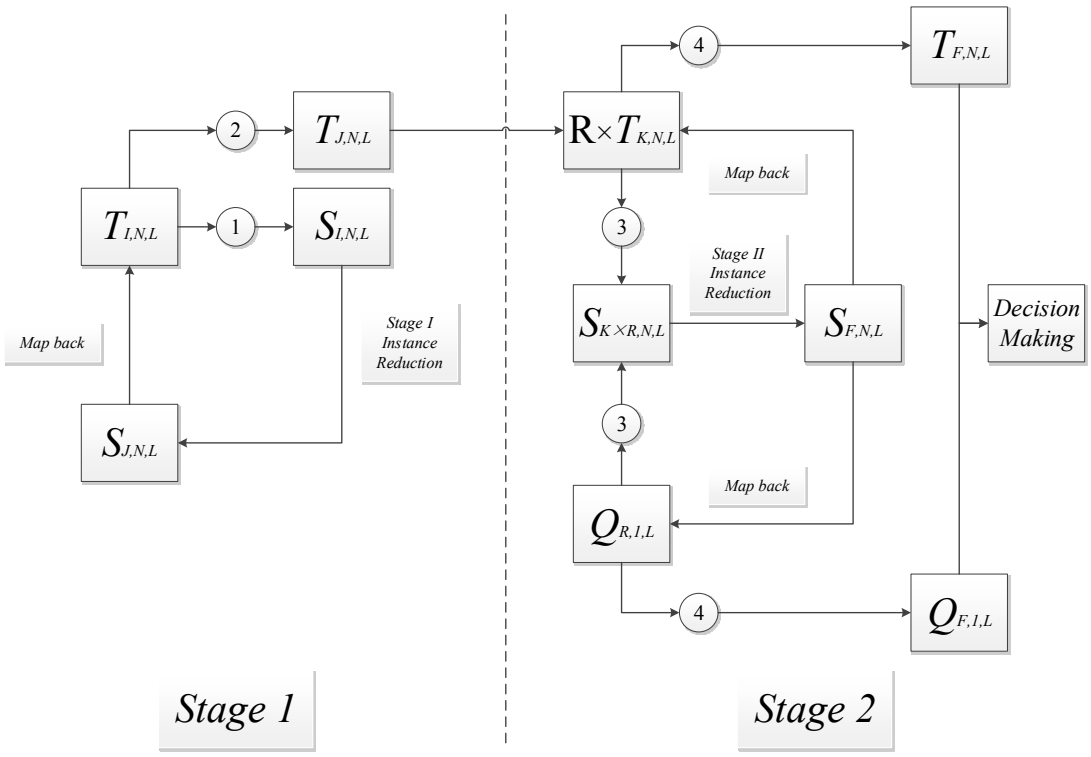


Fig. 8.2 Algorithm step-wised flow chart (start from $\boldsymbol{T_{I,N,L}}$, order: ① to ④), the purpose of this diagram is to illustrate the three main circles involved in the two stages.

Stage 1: Inter Class

For $i = [1, I]$, $n = [1, N]$, $l = [1, L]$,

1) Compute the distances: $d(T_{i,n,l}, T_{i,m,l})$, where m indicates a class, $m \neq n$.
2) Compute the sum of the resulting distances

$$d(T_{i,n,l}, T_{:,m,l}) = \sum_{i=1}^{i=I} d(T_{i,n,l}, T_{i,m,l})$$

where $m \neq n$, $T_{:,m,l}$ indicates all the elements of class m in dimension l.

3) Next compute the mean of the resulting distances of Step 2:

$$d(T_{i,n,l}, T_{:,\neg n,l}) = \frac{\sum_{i=1}^{n} d(T_{1,1,1}, T_{:,2,1}) + d(T_{1,1,1}, T_{:,3,1}) + \cdots + d(T_{1,1,1}, T_{:,i,1})}{n-1}$$

End the for-loop.

4) Use the resulting distances $d(T_{i,n,l}, T_{:,\neg n,l})$ to construct a distance score matrix $S_{I,N,L}$, which also has the same $I \times N \times L$ three dimensional structure as the matrix $T_{I,N,L}$. The distance scores in $S_{I,N,L}$ match with the location of their corresponding elements in the original training set matrix $T_{I,N,L}$.

For $i = [1, I]$, $n = [1, N]$, $l = [1, L]$,

5) For each dimension per class of $S_{I,N,L}$, rank the scores (I scores) in *descending* order and preserve the first J ($J < I$) scores with the highest values. Thus generate an ordered and reduced distance score set $S_{J,N,L}$. The choice of J depends on the database and the empirical analysis.

End the for-loop.

6) Utilize the scores in $S_{J,N,L}$ to map back to the original training set $T_{I,N,L}$ and select the corresponding J elements to form a new feature set $T_{J,N,L}$.

The resulting feature set $T_{J,N,L}$ contains the features with relatively large inter-class distances.

Stage 2: Intra Class

To alleviate the considerable template ageing efforts in EEG pattern recognition, *Stage 2* is developed for the algorithm. This stage involves the adaptive selection of the instances from both the training set $T_{J,N,L}$ and the query set $Q_{R,X,L}$, where R indicates the number of instances of the unknown identity query set X, the number of dimension of the feature vectors in Q should be the same as the training set T.

For $j = [1, J]$, $n = [1, N]$, $l = [1, L]$,

7) In $T_{J,N,L}$, divide the sorted and ranked J elements into c_1 segments (clusters): preserve $c_2$segment(s) ($c_2 < c_1$) which contain relatively high feature densities. Here the derivative/gradient of the feature element values may be used to measure the density of the pattern distributions. Thus, a subset of the feature set $T_{J,N,L}$ with denser clustered elements may be generated.

End the for-loop.

Denote the subset of $T_{J,N,L}$ as $T_{K,N,L}$ after Step 7 ($K \leq J$).

For $k = [1, K]$, $n = [1, N]$, $l = [1, L]$,

8) Measure the distances between the elements in training set and the elements in the query set $d(T_{k,n,l}, Q_{r,x,l})$, where $k \in [1, K], r \in [1, R], n \in [1, N], l \in [1, L], R \leq K$.

End the for-loop.

9) Since $R \leq K$, it may produce multiple distance score-matrixes $(K - R + 1) * S_{R,n,l}$ (while $R = K$ it produces only one score matrix $S_{R,n,l}$ per class n with R scores). Rank all the resulting score-matrixes per class per dimension in ascending order, i.e. concatenate $(K - R + 1)$ matrixes $S_{R,n,l}$ for ranking. Preserve the top F training-query element pairs with the *shortest* distances, $F \leq R$.
10) Facilitating the resulting score-matrixes $S_{F,n,l}$, where $n \in [1, N], l \in [1, L]$, map (trace) back to the training set $T_{K,N,L}$ and query set $Q_{R,X,L}$, for class n construct a training-query pair $(T_{F,n,L}, Q_{F,X,L})$.
11) Repeat 10) for each class of the training set $T_{K,N,L}$, therefore, N training-query pairs $(T_{F,n,L}, Q_{F,X,L})$ are adaptively generated for each class in the training set.
12) The distances between the instances of each specific query $Q_{f,X,l}$ and training set $T_{f,n,l}$ are measured and the membership voting scheme (here the standard nearest neighbour rule is employed) is used for the decision making.

8.2.2 Illustrative Example

For an illustrative example, consider a case where the training set T comprises three classes, i.e. N=3; the number of instance is 3, i.e. I=3; each instance contains two elements (two dimensions), i.e. L=2. Therefore, each element of the training set may be represented by a $\mathrm{I} \times \mathrm{N} \times \mathrm{L}, \mathrm{i.e.}\, 3 \times 3 \times 2$ matrix: $\mathrm{T}_{:,:,:} = \{\mathrm{T}_{:,:,1}, \mathrm{T}_{:,:,2}\} = \left\{\begin{bmatrix} 6 & 9 & 2 \\ 12 & 13 & 15 \\ 22 & 25 & 24 \end{bmatrix}, \begin{bmatrix} 5 & 6 & 7 \\ 17 & 14 & 19 \\ 21 & 25 & 27 \end{bmatrix}\right\}$, where, for example, $\mathrm{T}_{:,1,1} = [6\ 9\ 2]$, $\mathrm{T}_{1,1,1} = [6]$. The query set comprises a single class with unknown membership, and can be represented by a matrix $\mathrm{Q}_{:,\mathrm{x},:}$, where x indicates the unknown identity of the query. The query contains

two instances, i.e. I=2; dimension L=2. This example of query set may be illustrated with random numbers as $Q_{:,x,:} = \{Q_{:,x,1}, Q_{:,x,2}\} = \{[6\ 7], [5\ 8]\}$. It is noticed that the query contains fewer instances, which is the usual case. The algorithm is approximately illustrated following the same steps indicated in Section 3.2.

Stage 1

Step 1 Compute the Euclidean distances d between classes in the training set.

$d(T_{1,1,1}, T_{1,2,1}) = d(6, 12) \approx 13, d(T_{1,1,1}, T_{1,3,1}) = d(6, 22) \approx 23;$

$d(T_{1,1,1}, T_{2,2,1}) = d(6, 13) \approx 14, d(T_{1,1,1}, T_{2,3,1}) = d(6, 25) \approx 26;$

$d(T_{1,1,1}, T_{3,2,1}) = d(6, 15) \approx 16, d(T_{1,1,1}, T_{3,3,1}) = d(6, 24) \approx 25;$

Step 2 Sum above-computed scores: $d(T_{1,1,1}, T_{:,2,1}) = 43; d(T_{1,1,1}, T_{:,3,1}) = 74.$

Step 3 Compute the mean of $d(T_{1,1,1}, T_{:,2,1})$ and $d(T_{1,1,1}, T_{:,3,1})$: $d(T_{1,1,1}, T_{:,:,1}) = \frac{d(T_{1,1,1}, T_{:,2,1}) + d(T_{1,1,1}, T_{:,3,1})}{2} \approx 59.$

Step 4 Repeat Step 1-Step 3 for I=3, N=2, L=2. Therefore, an inter-class distance score matrix indicates each elements of $T_{:,:,:}$ is generated $S_{:,:,:} = \{S_{:,:,1}, S_{:,:,2}\} \approx \left\{\begin{bmatrix} 59 & 63 & 56 \\ 60 & 62 & 66 \\ 74 & 82 & 79 \end{bmatrix}, \begin{bmatrix} 64 & 65 & 63 \\ 72 & 65 & 76 \\ 73 & 84 & 89 \end{bmatrix}\right\}.$

Step 5 Rank the scores of $S_{:,:,:}$ per class per dimension in descending order, hence construct a new score matrix $S'_{:,:,:} = \{S'_{:,:,1}, S'_{:,:,2}\} \approx \left\{\begin{bmatrix} 63 & 59 & 56 \\ 66 & 62 & 60 \\ 82 & 79 & 74 \end{bmatrix}, \begin{bmatrix} 65 & 64 & 63 \\ 76 & 72 & 65 \\ 89 & 84 & 73 \end{bmatrix}\right\}.$

Step 6 Preserve the first two instances per class per dimension of $S'_{:,:,:}$, hence $S''_{:,:,:} = \{S''_{:,:,1}, S''_{:,:,2}\} \approx \left\{\begin{bmatrix} 63 & 59 \\ 66 & 62 \\ 82 & 79 \end{bmatrix}, \begin{bmatrix} 65 & 64 \\ 76 & 72 \\ 89 & 84 \end{bmatrix}\right\}.$

Step 7 Map the scores of $S''_{:,:,:}$ back to their corresponding elements in $T_{:,:,:}$. Hence, the reconstructed training set $T'_{:,:,:} = \left\{\begin{bmatrix} 9 & 6 \\ 15 & 13 \\ 25 & 24 \end{bmatrix}, \begin{bmatrix} 6 & 5 \\ 19 & 17 \\ 27 & 25 \end{bmatrix}\right\}$, where the preserved number of instances per class per dimension is J=2.

Stage 2

Step 8 Compute the Euclidean distances d between Class 1 $T'_{:,1,:} = \{[9\ 6], [6\ 5]\}$ and query set $Q_{:,x,:} = \{Q_{:,x,1}, Q_{:,x,2}\} = \{[6\ 7], [5\ 8]\}$.

$d(T'_{1,1,1}, Q_{1,x,1}) = d(9,6) \approx 10.8, d(T'_{1,1,1}, Q_{2,x,1}) = d(9,7) \approx 11.4;$

$d(T'_{2,1,1}, Q_{1,x,1}) = d(6,6) \approx 8.5, d(T'_{2,1,1}, Q_{2,x,1}) = d(6,7) \approx 9.2;$

$d(T'_{1,1,2}, Q_{1,x,2}) = d(6,5) \approx 7.8, d(T'_{1,1,2}, Q_{2,x,2}) = d(6,8) \approx 10.0;$

$d(T'_{2,1,2}, Q_{1,x,2}) = d(5,5) \approx 7.1, d(T'_{2,1,2}, Q_{2,x,2}) = d(5,8) \approx 9.4.$

Step 9 Generate the intra-class distance score matrix for training set of Class 1 and Query set pair: $S_{c1} = \left\{ \begin{bmatrix} d(T'_{1,1,1}, Q_{1,x,1}) \\ d(T'_{2,1,1}, Q_{1,x,1}) \\ d(T'_{1,1,1}, Q_{2,x,1}) \\ d(T'_{2,1,1}, Q_{2,x,1}) \end{bmatrix}, \begin{bmatrix} d(T'_{1,1,2}, Q_{1,x,2}) \\ d(T'_{2,1,2}, Q_{1,x,2}) \\ d(T'_{1,1,2}, Q_{2,x,2}) \\ d(T'_{2,1,2}, Q_{2,x,2}) \end{bmatrix} \right\} = \left\{ \begin{bmatrix} 10.8 \\ 8.5 \\ 11.4 \\ 9.2 \end{bmatrix}, \begin{bmatrix} 7.8 \\ 7.1 \\ 10.0 \\ 9.4 \end{bmatrix} \right\}.$

Step 10 Rank the scores of S_{c1} per dimension in ascending order hence construct a new score matrix $S'_{c1} = \{[8.5\ 9.2\ 10.8\ 11.4], [7.1\ 7.8\ 9.4\ 10.0]\}$.

Step 11 Preserve the first instance of S'_{c1} for each dimension, hence , $S''_{c1} = \{[d(T'_{2,1,1}, Q_{1,x,1})], [d(T'_{2,1,2}, Q_{1,x,2})]\} = \{[8.5], [7.1]\}$.

Step 12 Map the scores of S''_{c1} back to their corresponding elements in $T'_{:,1,:}$ and $Q_{:,x,:}$. Hence the reconstructed training set of Class 1 $T''_{:,1,:} = \{[6], [5]\}$, which is the best subset of the query subset $Q'_{:,x,:} = \{[6], [5]\}$ in this case.

Step 13 Repeat Step 8 to Step 12 for all the available classes in $T'_{:,:,:}$. It is found $T''_{:,2,:} = \{[13], [17]\}$; $T''_{:,3,:} = \{[24], [25]\}$. The reconstructed training set is $T''_{:,:,:} = \left\{ \begin{bmatrix} 6 \\ 13 \\ 24 \end{bmatrix}, \begin{bmatrix} 5 \\ 17 \\ 25 \end{bmatrix} \right\}$ the query subset is $Q'_{:,x,:} = \{[6], [5]\}$.

Step 14 Membership of the query is classified by comparing the instance similarity (absolute value) between $Q'_{:,x,:}$ and $T''_{:,N,:}$ (N=3), and allocated to Class 1.

Figure 8.3 visualises the proposed algorithm using the illustrative example. Three classes, with each contain three instances; each instance is made up of a two dimensional scalar.

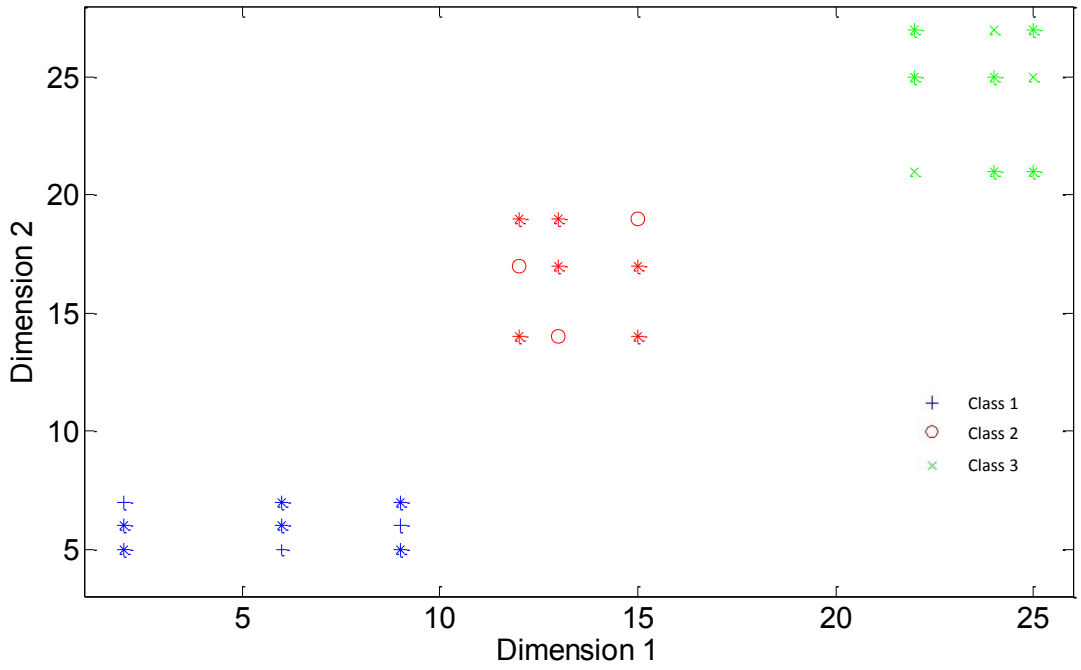


Fig. 8.3 Visualization of the illustrative example using the proposed algorithm

The instances of the three classes involved in the illustrative example are depicted in Fig. 8.3 with three different shapes. The stars indicate the reconstructed vectors (templates). Based on the available measurements, a series of new templates are constructed for each class in place of the original feature vector. The proposed algorithm offers the adaptability to optimize the inter-class (*Stage 1*) and intra-class (*Stage 2*) modelling using the obtained measurement: all the features used to form new templates are from the original data, but better organized for each dimension.

8.2.3 Remarks

Some remarks are presented in this section for better clarification of the proposed algorithm. The "Step" mentioned in the remarks relates to the step label in Section 8.2.1.

Remark 1: Step 1) of the proposed algorithm shares a certain similarity with the initial step for computing the Hausdorff Distance [177]. Instead of measuring the distance between feature vectors, the distance between elements of the feature vector from different classes are measured.

Remark 2: The preserved instances in $T_{J,N,L}$ ($J < I$) have better inter-classes separation. This is achieved by 1) removing the elements (which form the instances) with relatively short inter-class distances and 2) using the preserved elements to construct new instances (templates).

Remark 3: The training set $T_{J,N,L}$ comprises a series of new feature vectors (templates/representations), these feature vector template(s) may be reconstructed using the elements from different feature vector templates of the original set $T_{I,N,L}$, i.e. the templates of the subset $T_{J,N,L}$ may be a series of newly constructed (inter-classes optimised) instances (using the available elements) which may not be found their correspondences in the original set as feature vectors.

Remark 4: The reconstructed templates/instances are generated so as to maximize the inter-classes distances for every single dimension. *Stage 2* is designed to minimize the intra-class distances by generating a subset of $T_{F,N,L}$ for each dimension, given the elements of each class in the feature set $T_{K,N,L}$ has an improved inter-classes separation (in *Stage 1*).

Remark 5: Step 7 is optional in this algorithm, as it is designed to remove the possible noise/outlier points. Depending on the quality of the data (signals or images), it may not help in improving recognition performance.

Remark 6: In Step 10, for each class an intra-class distance-minimized training-query pair $T_{F,n,L}$ and $Q_{F,X,L}$ for the particular query set $Q_{R,X,L}$ are generated. For the verification scenario, the nearest instance selection in Step 10 only operates on one class as opposed to all the classes for the identification scenario.

Remark 7: Since the selection of the optimal elements in the training set $T_{K,N,L}$ is performed for each dimension independently (L dimensions, same as *Stage 1*): the instances of the reduced training set $T_{F,N,L}$ are a series of reconstructed new feature vectors too, which possess the highest intra class similarity. The elements of the newly constructed templates are the selectively assembled forms that stem from all the available templates, i.e. the final feature vectors in $T_{F,N,L}$ may have no correspondences in $T_{K,N,L}$ as feature vectors.

Remark 8: The final number of instances F usually is very small: as a rule of thumb, the number of F is set to be the same as the number of instances in the instance-reduced query set $Q_{F,X,L}$ for the optimal training-query matching pair.

Remark 9: *Stage 1* alone can be viewed as an independent algorithm for pattern recognition with improved inter-class separation. *Stage 2* may be applied for some image-based applications, but may not be equally effective for recognition of time-series data (such as EEG signals and voice) without performing *Stage 1* first. Two reasons: 1) intra-class similarity is low in many such signals and 2) inter-class patterns overlap in the feature space (would lead to high FAR).

Remark 10: The three main characteristics of the proposed algorithm: 1) the resulting templates of *Stage 1* and *Stage 2* may be new templates, which have less inter-classes similarities (after *Stage 1*) and shorter intra-class distances (after *Stage 2*); 2) Stage 1 of the algorithm constructs new hyper-boundaries by removing less discriminating instances (even though they may not be due to noise) to achieve better inter-class separation; 3) the *Stage 2* algorithm performs an adaptive learning process depending on each specific query to enhance the intra-class similarity, and significantly reduces the volume of the final training set $T_{F,N,L}$. In one of the case studies in Section 8.4, the proposed algorithm preserved only 1.7% of the training set: 4 out of 240 instances were kept ($F = 4$), tested using two EEG databases.

Remark 11: Figure 8.4 shows the Instance-based Template Reconstruction Learning (ITR) algorithm in a naïve way. As an example, assuming a single class feature set with three instances (observations), each instance contains a five dimensional feature vector.

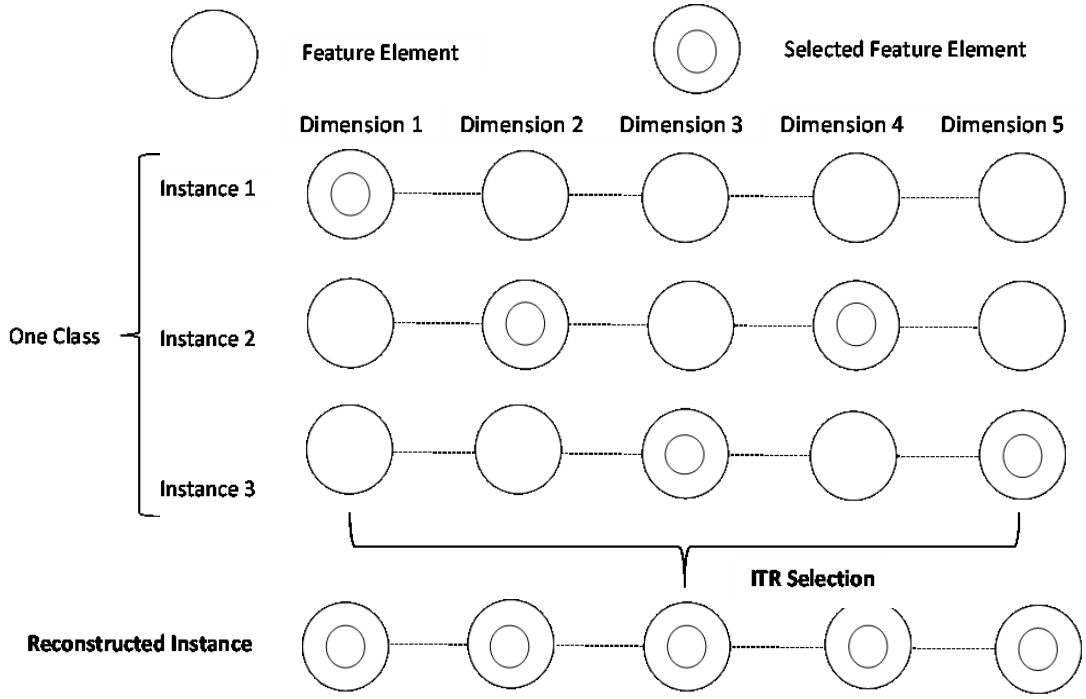


Fig. 8.4 Naïve illustration of the ITR algorithm

By applying the ITR algorithm, the elements of the three instances are extensively measured and one new instance is constructed based on the conditions of *Stage 1* and *Stage 2*. It is clear the selecting and reconstructing of the instance elements process is dimension-independent, i.e. the process is performed dimension by dimension, the order of feature dimensions for each feature vector is not changed (horizontal, Fig. 8.4), only the order of the instances (vertical, Fig. 8.4) may change depend on the proposed algorithm.

8.3 Performance Evaluation

This section illustrates the performance of the proposed instance-based template reconstruction learning algorithm by a binary classification problem using EEG data. The algorithm is then compared with two other conventional instance-based learning algorithms: *k*-NN and SVM. As one of the most popular instance-based learning algorithm, SVM has been proved as an effective and efficient learning algorithm in solving many real world problems, such as classifying proteins in medical science, classification of images and hand-written characters [178][179][180]. MM/I database

and Mobile Sensor Database are used to test and compare the performances of *k*-NN, SVM and ITR algorithm.

8.3.1 Efficacy of the Algorithm

The data of the first two subjects (S1-S2) in the Mobile Sensor Database were employed for this experiment. The data collection settings for the Mobile Sensor Database is much more challenging than that for some of the other publicly available databases used for biometric system evaluation (such as MM/I dataset and VEP database). This may be better in revealing the effectiveness of ITR algorithm when dealing with more difficult datasets. Instead of HHT-based features which provided the best performance amongst the proposed new features (see Chapter 6), a wavelet-based feature was employed for this two-class problem because of its ease of computation [108].

Fig. 8.5 shows the original pattern distribution of the first two dimensions in the feature vectors (there are five dimensions in total for each feature vector). The graph shows the distribution of extracted training set instances: each class contains 240 instances (points in the figure). C1 and C2 indicate the patterns of the two classes. It can be observed that the patterns of these two classes extracted from the low quality raw EEG data are quite intertwined in the feature space. Following the definitions in Section 8.2.1, this two-classes training set can be represented as $T_{I,N,L} = T_{240,2,5}$. *Stage 1* of the ITR algorithm is designed for improving inter-class discrimination of the training set by reducing the number of instances (I) for each class.

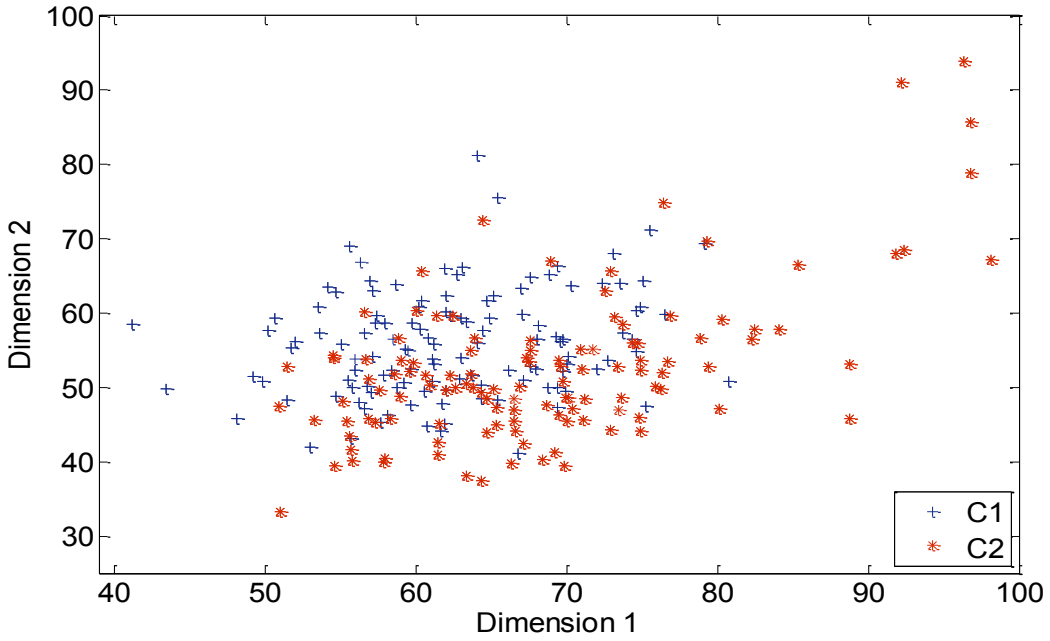


Fig. 8.5 Original feature patterns

Fig. 8.6 depicts the resulting patterns after performing *Stage 1* of the ITR algorithm. The instance patterns of two classes are better separated, only 3/5 of all the elements with the largest inter-class distances are preserved. The preserved training set is denoted as $T_{j,n,l} = T_{144,2,5}$. For comparison purposes, the axis ranges of the depicted feature space in the following graphs (Fig. 8.6 to Fig. 8.8) are kept the same as in Fig. 8.4 and Fig. 8.5.

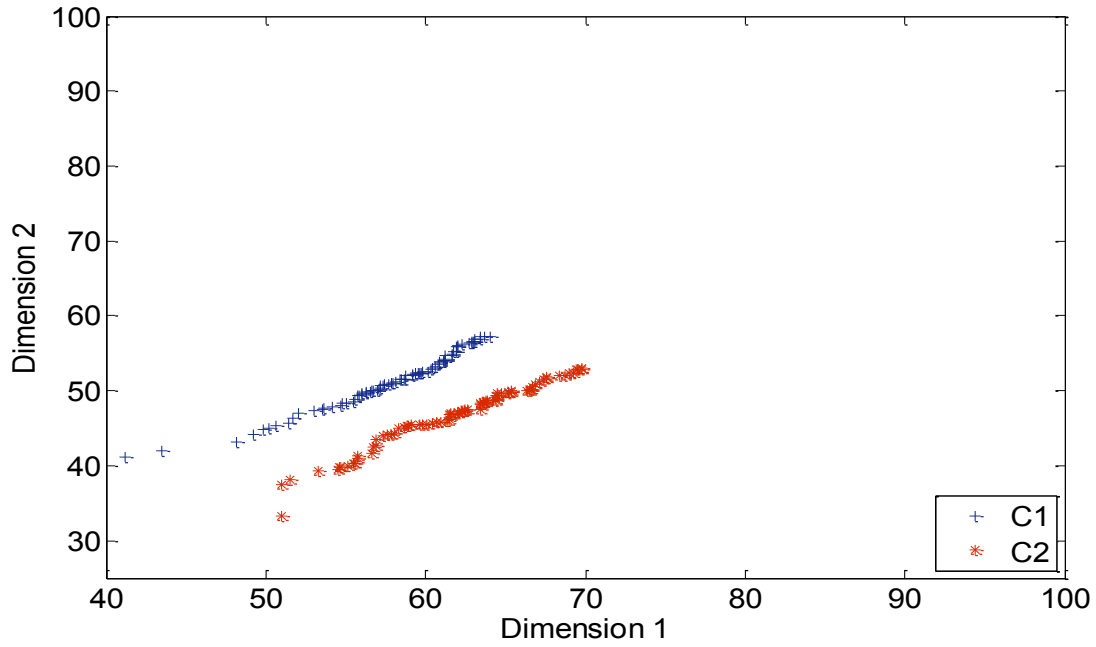


Fig. 8.6 New training set patterns after *Stage 1*

Fig. 8.7 illustrates the effectiveness of the pattern distribution after performing the density-based instance reduction (Step 7 of the proposed algorithm): only condensed pattern clusters are preserved. As was mentioned in Section 8.2, Step 7 of the algorithm is an optional operation depending on the characteristics of the employed data. The number of the preserved instance patterns in Fig. 8.7 is further reduced to K=72, which are only half the amount of the previous step (J=144): now the training set is a new subset with more condensed instances $T_{K,N,L} = T_{72,2,5}$.

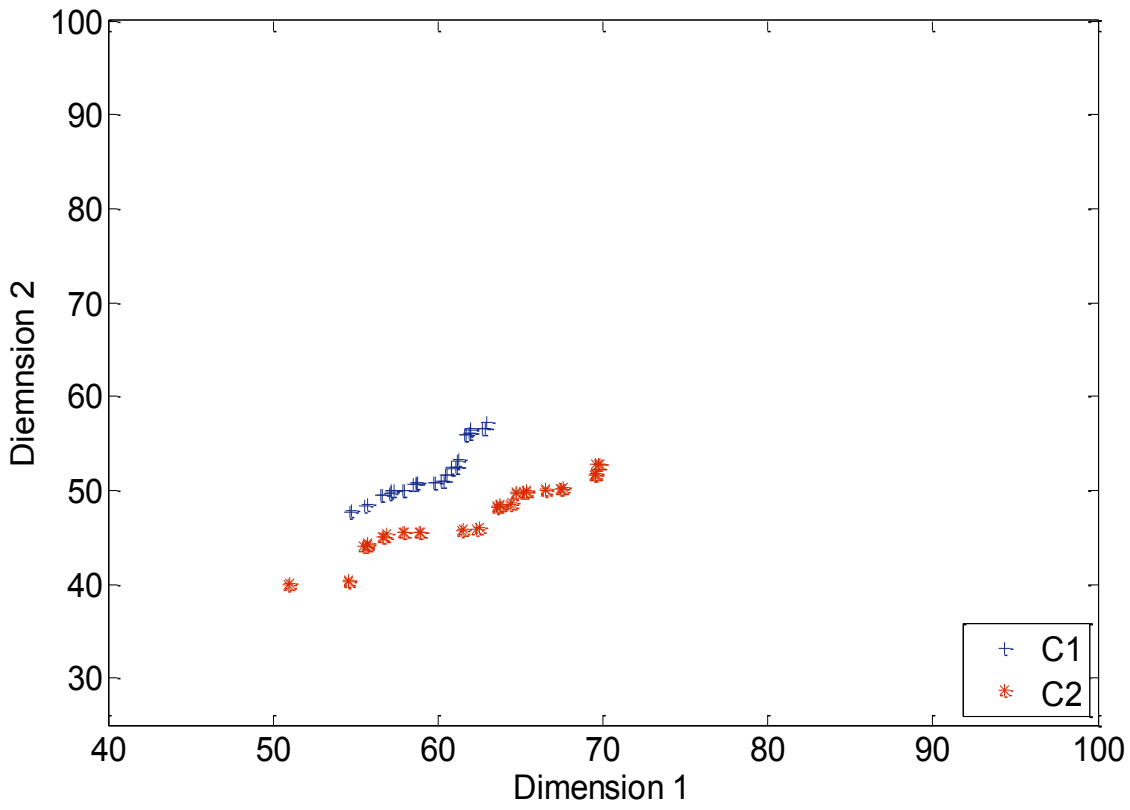


Fig. 8.7 Training set after further density-based instance reduction

It is worth to noting that, since the distributions of the instances after *Stage 1* may not be overlapping (Fig. 8.6 and Fig. 8.7), the classification performance of the proposed ITR algorithm is no longer bounded by the Bayes Error Rate [181].

In order to maximize the usage of the preserved instances, the next step of the proposed algorithm is designed to minimize the intra-class distance between the training set and the query set. Fig. 8.8 shows the resulting classification boundary for a two-class query attempt. The algorithm is designed to select the best templates from the reduced training set $T_{k,n,l}$ and construct a further reduced subset $T_{f,n,l}$ before final decision making. In this study, the 1-NN classifier was employed for generating the decision boundary. *Stage 2* of the ITR algorithm was applied to the test set in order to enhance the quality of the query set as well.

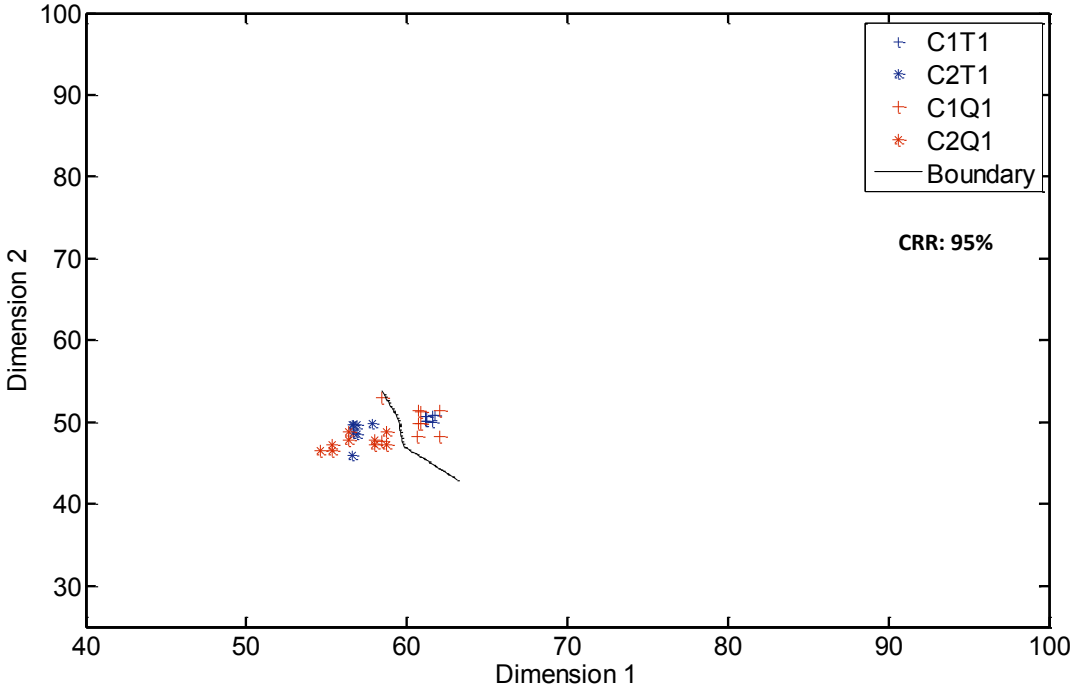


Fig. 8.8 Binary classification for query 1

As the result of the *Stage 2*'s operations, only the best 10 instances (templates) was adaptively selected for both the training set and the query set, i.e. the final sets are $T_{f,n,l} = T_{10,2,5}$ and $Q_{f,x,l} = Q_{10,x,5}$ for this binary classification.

As an adaptive learning algorithm, the selection of the subset (for training) shall depend on the property of each particular query: it gives the best reconstructed templates each class has to offer from the training set for that query to maximize the intra-class similarity. Therefore, the template ageing effect (especially for EEG signals) may be alleviated.

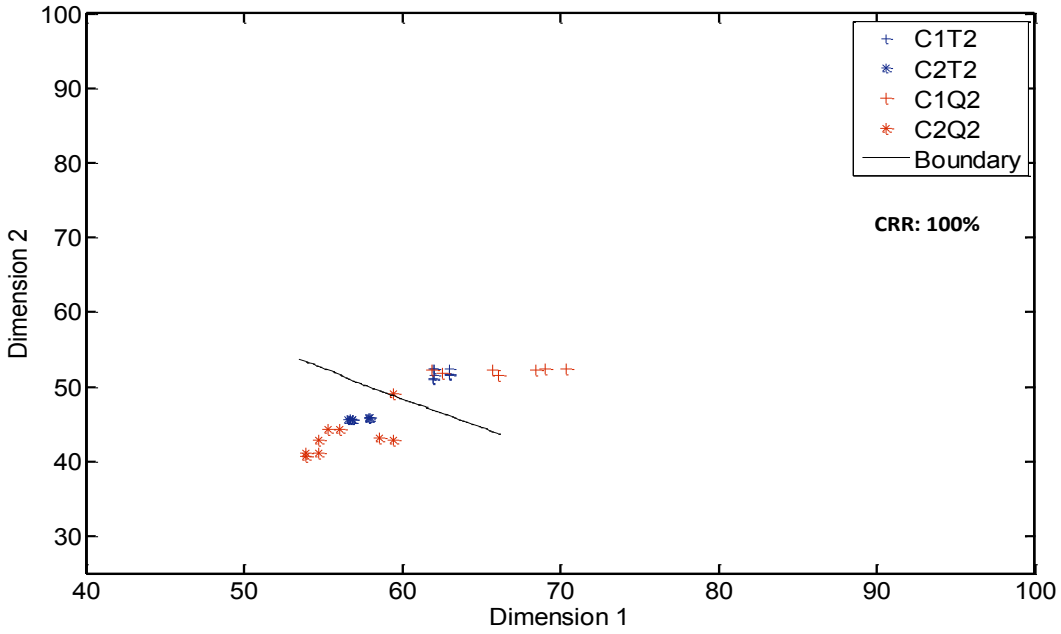


Fig. 8.9 Binary classification for query 2

As an example, for a different two-classes query (Q2, another 10 seconds' recording), a different decision boundary in Fig. 8.9 (compared with Fig. 8.8) indicates a series of different instances (f =10) from the training set and query set are selected for classification. The query set which achieved 95% of CRR in this investigation is a 10 seconds' EEG recording from a separated session, whereas the classification performance using the original all instance patterns of the same 10 seconds' data (illustrated by Fig. 8.5) provided accuracy of only 37.5% using a 1-NN classifier.

8.3.2 Comparison with Relevant Learning Algorithms

The proposed method is an instance-based learning algorithm; k-NN and SVM [182] are two popular algorithms which belong to this category as well. In this section some comparisons between the proposed algorithm and these two algorithms are provided. Two important properties may affect the performance of learning algorithms in pattern classification: 1) the relative locations of feature vectors and, 2) the feature space for decision boundary generation.

The k-NN algorithm and its variants achieve class separation by finding the feature vectors (instances, or templates) with optimal geometric locations. By removing less

significant instances, instance-reduced *k*-NN trims the distribution of instances and forms a more class-discriminating training model. It must be clarified that the "neighbour" (or instance) in *k*-NN is a feature vector which normally contains multiple elements (components): each element represents one dimension in the multidimensional space and the distance is measured between feature vectors. Therefore, the *k*-NN algorithm may perform poorly when applied to high dimensional instances as most of the instance vectors with high dimensionality in the training set (from different classes) may be almost equidistant to the query vector [183].

The original SVM is designed to utilize the so-called supporting vectors to achieve class separation in the original Euclidean hyper planes. For the linear SVM algorithm, only the supporting vectors (instances) are considered during the classification. The kernel-SVM algorithm (nonlinear SVM), however, employs the so-called kernel function to transform the patterns in the original feature space into a much higher dimensional space. The motivation of performing such a trick is due to the fact that the feature vectors (instances) in the original space are often linearly inseparable: by mapping the feature vectors into kernel space (which potentially can be an infinite dimensional space), it may be possible to find the distinctive supporting hyper-boundaries between two classes in that new space. As opposed to proposed algorithm as well as *k*-NN algorithms, which need to precisely locate each feature (by measuring the distance, for instance), for SVM only the results of kernel mapping are required for pattern classification in the kernel space [182].

Although the ITR algorithm could generally be fitted into the *k*-NN framework, it also contains a feature extraction aspect: The standard *k*-NN selects the instances with better geometric locations. The proposed instance-based template reconstruction algorithm optimizes the instances in even lower level: it selects the elements (not vectors) with the best locations for every instance per dimension, using the optimal elements to construct new (and better) instances (feature vectors). Therefore, the resulting instance vectors are most likely different from any of the original instance candidates (though statistically it is possible that the optimal instance with optimal elements may exist in the original training set). Furthermore, by referring to Fig. 9.1,

such a scheme also optimizes each feature vector dimension by dimension, but without the dimension number reduction (unlike the Principal Component Analysis).

ITR algorithm with the elemental level analysis provides more flexibility in enhancing the inter-class separation (*Stage 1*) and intra-class similarity (*Stage 2*): in *Stage 1* the algorithm creates a series of new hyper-planes formed by the reconstructed instance vectors to achieve better between-class discrimination; in *Stage 2*, for each different query instance, a series of new hyper-planes for each class is constructed to minimize the within-class variation.

8.4 Experimental Case Studies

Three case studies are provided in this section to evaluate the effectiveness of the ITR algorithm. In Section 8.4.1, the MM/I dataset is used for motor movement/imagery task classification: the aim is to solve a four-class recognition problem. The results are generalized using a leave-one-out three-fold cross-validation scheme: the data of each single run forms one fold during the validating process. Section 8.4.2 presents the evaluation results for biometric recognition: using the data of 105 selected subjects from MM/I dataset, the aim is to classify the instances from 105 classes. In Section 8.4.3, the Mobile Sensor Database is used to investigate the effectiveness of the ITR algorithm in exploiting the small volume of low quality training data that is available for biometric recognition. The EEG data obtained from 27 subjects (classes) was used, of each one minute's recording per session (two sessions in all), 10 seconds' data was used as the query set and the rest was used for training the system. The impact of the template ageing effects is also explored in the last set of experiments. The overall system diagram is illustrated in Fig. 8.10.

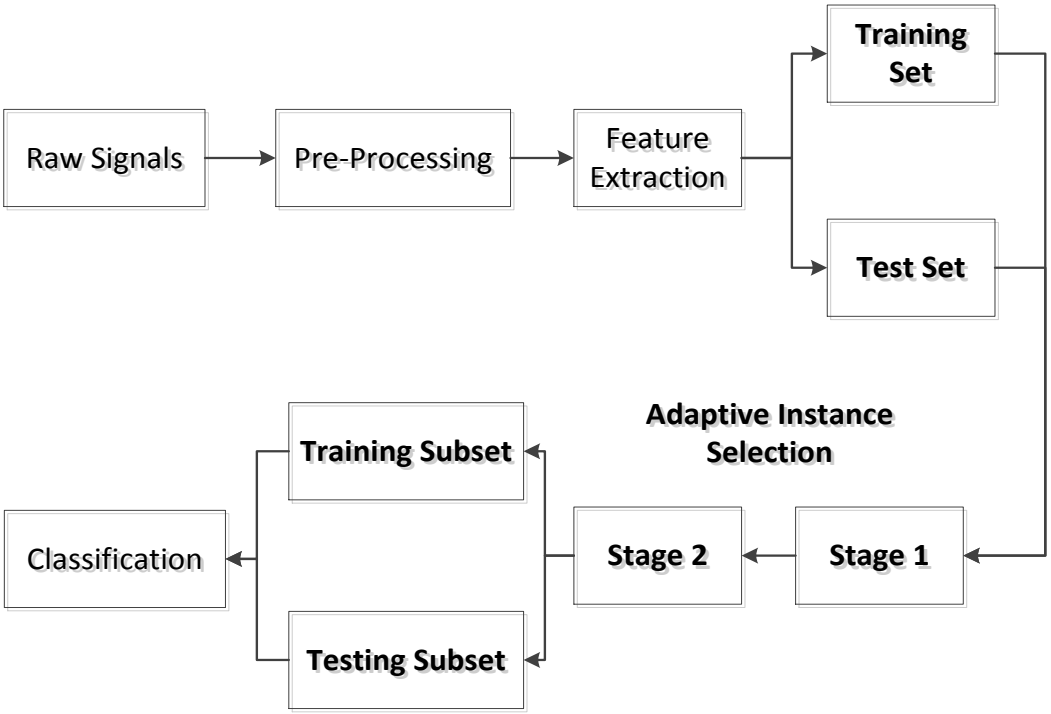


Fig. 8.10 Overall system diagram

8.4.1 Classification of Motor Movement/Imagery Tasks

One of the early studies in mental tasks classification was conducted by Anderson *et al*. [184], they reported an average accuracy of 91.4% for a two class problem. Multiple features were investigated and the Autoregressive (AR) coefficients provided both the best and the most consistent performance. Data of four subjects were used in their experiment; the collected EEG data (using six electrodes) contained two sessions, each session included 10 seconds' recording and repeated 5 times per session. 80% of the data were used for training an Artificial Neural Network (ANN) classifier. By employing the same database, Palaniappan [185] reported a Correct Recognition Rate of 97.5%; the proposed feature was the spectral powers of four typical EEG bands.

Akrami *et al*. [186] further investigated the performance of classifying three mental tasks. Data was obtained from one left-handed male subject; three mental imagery tasks related to hand movements were performed during the experiment. A classification rate of 95.71% was obtained by employing only 2 electrodes (C3 and

C4) for data collection. The bandwidth specified logarithmic power spectral density of the signals was used as feature; the ANN classifier provided the highest accuracy.

However, the performance seems to decrease significantly when the number of classes is increased. Li *et al*. [187] reported their performances on the classification of five mental tasks. The experiment involved two subjects and the classification rate of 76.3% was reported. The entropy of wavelet coefficients was used as features and a Support Vector Machine classifier using the Radial Basis kernel Function (RBF) was used for classification. Using only three classes from the same database, Abdalsalam *et al*. [188] achieved an improved accuracy of 80.4%, with wavelet transform for feature extraction.

To alleviate the above-mentioned performance degradation for classification while the number of mental classes was increased, the Empirical Mode Decomposition (EMD) for feature extraction was used by Diez *et al*. [189] who reported an average classification rate of 91% for five mental tasks. They used a database of 7 subjects and computed six different features using the resulting Intrinsic Mode Functions (IMFs) after EMD.

8.4.1.1 Database and Overall Experimental Design

In this work, the wavelet-based feature is used for testing the proposed algorithm. The MM/I dataset is employed in this investigation for classifying four motor movement/imagery tasks.

Three recording runs were made for each mental task, and each run lasted for about 2 minutes. This public database has been used for task classification and some of the relevant literature is briefly reviewed here to facilitate comparison with the proposed technique. Sleight *et al*. [190] used the data of 103 selected subjects out of 109; data from all the 64 electrodes were used for processing. The averaged power of multiple frequency bands were computed and used as features, which were then fed into a Gaussian-kernel SVM for classification. Their experiment was designed to classify between only movement and imagery tasks: the obtained highest accuracy was 69%. Tolić and Jović [191] also investigated the classification of these two types of mental

tasks using this database. A wavelet-based feature extraction method was used. Data of four subjects and only three electrodes were used and the best average accuracy was 68.21%. Loboda *et al*. [192] further investigated the classification of the four classes of different mental tasks using this database. Their reported highest correct discriminating rate was 82.5%, which was obtained by analysing the beta band's data from three electrodes.

In this case study, only the data from the Oz electrode was employed for easy-deployment and fast processing. The experimental chain is briefly stated as below:

1) The raw EEG signals of four mental tasks (T1-T4) are segmented into multiple windows, each window lasts for 4 seconds. Each window overlaps its adjacent window (50% overlapping).
2) The Wavelet Transform [59] is computed for each time-domain window, the wavelet decomposition is performed up to level 3 (eight bands are generated).

The 0Hz-50Hz band of the resulting wavelet coefficients were preserved for feature extraction; every generated feature vector contains five dimensions (so each dimension corresponding to a bandwidth of 10Hz).

3) The variance of the preserved wavelet coefficients in each window was computed and used as feature.
4) The proposed instance selection algorithm was applied to enhance the feature model construction.
5) The selected features were fed into a 1-Nearest Neighbour (1-NN) classifier to evaluate the task classification precision.

Data of the first fifteen subjects (S1-S15) of MM/I dataset were used for task classification: the main goal of this work is to investigate the classification of the mental tasks; multiple subjects are employed to estimate the variance of the overall performance. Data of two runs were employed for generating the prototypes (four minutes in all); the remaining data was used as the query.

8.4.1.2 Algorithm Comparison

The performance comparison was made by employing the standard 1-NN classifier, the non-linear SVM (second order polynomial kernel function) and the ITR algorithm. Fig. 8.11 illustrates the two dimensional distribution of the prototype patterns and test query, using the data of S1. The four different shapes/symbols indicate four different tasks; the blue shapes indicate the prototypes and the red shapes indicate the query points. For example, T1P indicates the prototype patterns of Task 1 and T1Q indicates the query patterns of Task 1. Fig. 8.12 and Fig. 8.13 also follow these abbreviations.

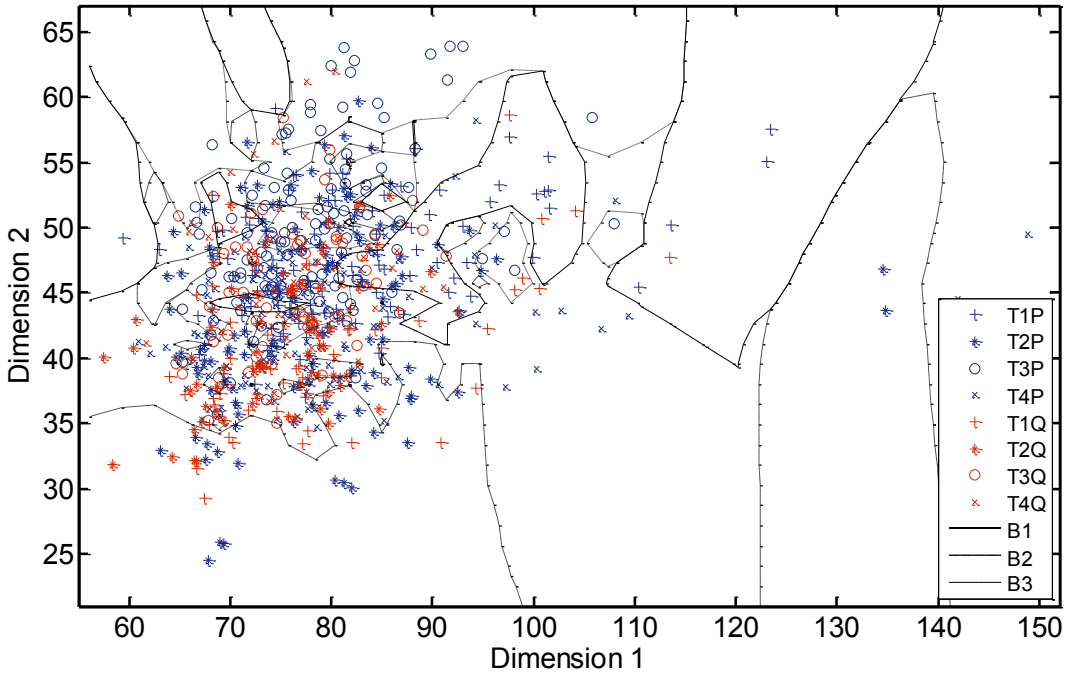


Fig. 8.11 Classification boundaries of 1-NN for S1

Fig. 8.11 shows the classification precision and decision boundaries when using the 1-NN algorithm: the resulting recognition accuracy of 21.25% is worse than random guess (25%). It has been shown that changing the window size can affect performance substantially [108]. Increasing the window size from 4 seconds (640 samples) to 40 seconds (6400 samples) the accuracy improved to 62.5%. However, in order to achieve this recognition rate, the restriction is that each feature must be computed from a recording window of 40 seconds.

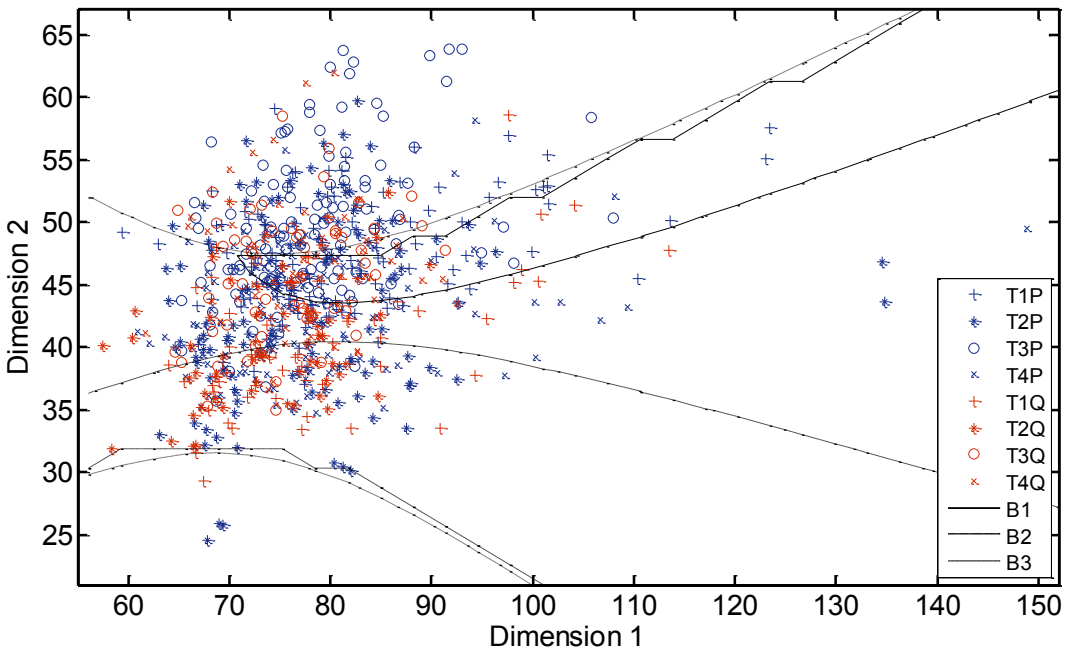


Fig. 8.12 Classification boundaries of SVM for S1

Fig. 8.12 indicates the decision boundaries for S1 constructed using a nonlinear SVM classifier with second order polynomial kernel function. Given the same pattern distribution (and same feature extraction method) as illustrated in Fig. 8.9 for both prototypes and queries, the SVM provided a better class separation with CRR of 33.33% (higher than 25%). As it is clear in both Fig. 8.11 and Fig. 8.12, the original four-class instance patterns are quite intertwined in the feature space, posing a very challenging task for any machine-learning algorithm.

The proposed algorithm is designed to prune the pattern distribution by removing the less discriminating instances and constructing new instances by combining the existing data elements. After applying the ITR algorithm, ideally the reshaped new prototype model will provide better class separation (*Stage 1*), and the adaptive learning scheme for each new query is designed to reduce the intra-class distribution variance (*Stage 2*).

Fig. 9.13 shows the pattern distribution and the recognition precision of 95% using the data from S1 after applying the proposed ITR algorithm. The ITR algorithm removes and reconstructs new templates for training: for any particular query only a

small amount of selected instances (multiple dimensions) in training set per class are preserved, which provide the highest intra-class similarity to that query.

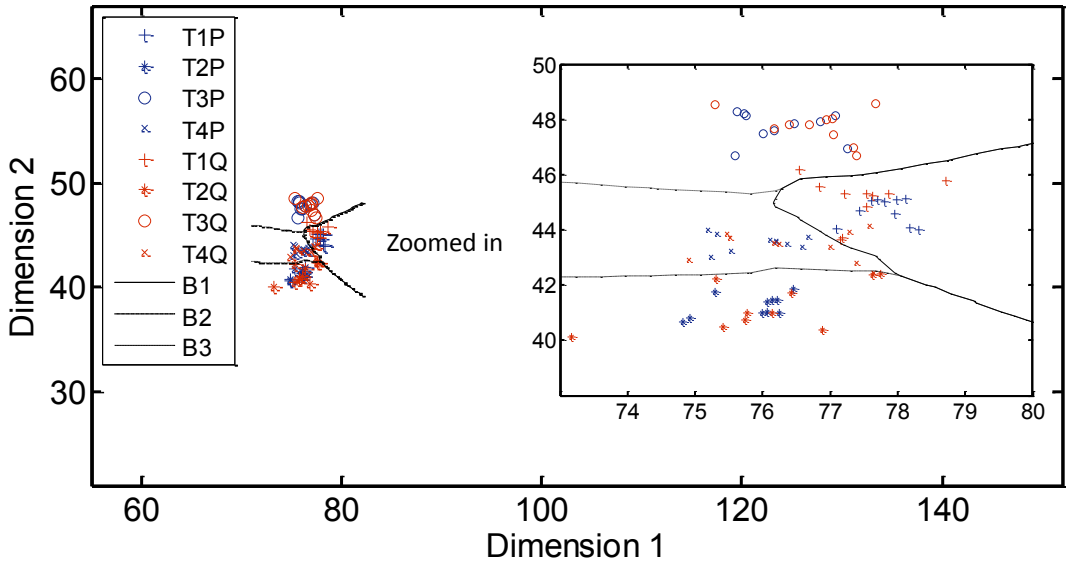


Fig. 8.13 Classification boundaries of ITR algorithm for S1

To further explain the feature points in Fig. 8.13, out of 120 prototypes per class ($I = 120$), only 10 selected prototypes ($F = 10$) were reconstructed using the elements from all the available prototypes, which are indicated by the blue points of four different shapes per class in the plot. The patterns in the query set were also modelled using ITR algorithm: $N = 4, L = 5$, the red dots in the graph indicate their relative geometric locations to the prototype patterns (blue dots). Comparing the cluster ranges (along the Dimension 1 and Dimension 2 axis) between Fig. 8.11 and Fig. 8.12, it can be seen that only the centre regions of the original clusters are presented and the patterns of interest are well distinguished.

8.4.1.3 Effectiveness for 15 Subjects

The overall results of motor movement/imagery tasks classification using the ITR algorithm incorporating the 1-NN algorithm are illustrated in Fig. 8.14. EEG data of the first 15 subjects (S1-S15) were used to evaluate the algorithms: due to the individual differences, feature vectors of different subjects were not concatenated and the task classification was evaluated per individual. In other report for task

classification [190], the data of same task from multiple individuals is concatenated into a single array for analysis.

The averaged accuracy of the 15 subjects using the 1-NN algorithm with only the Oz electrode data is around 62%, which is a comparable accuracy to the start-of-the-art reports [190][191][192]. The ITR algorithm incorporating the 1-NN algorithm provided a much better recognition performance, and only the mental task data for three subjects (S1, S7 and S10) were not 100% classified out of 15 subjects. The results in Fig. 8.14 indicate the ITR provided a better (average recognition rate of 98.38% and a lower performance variance compared with using the standard 1-NN algorithm.

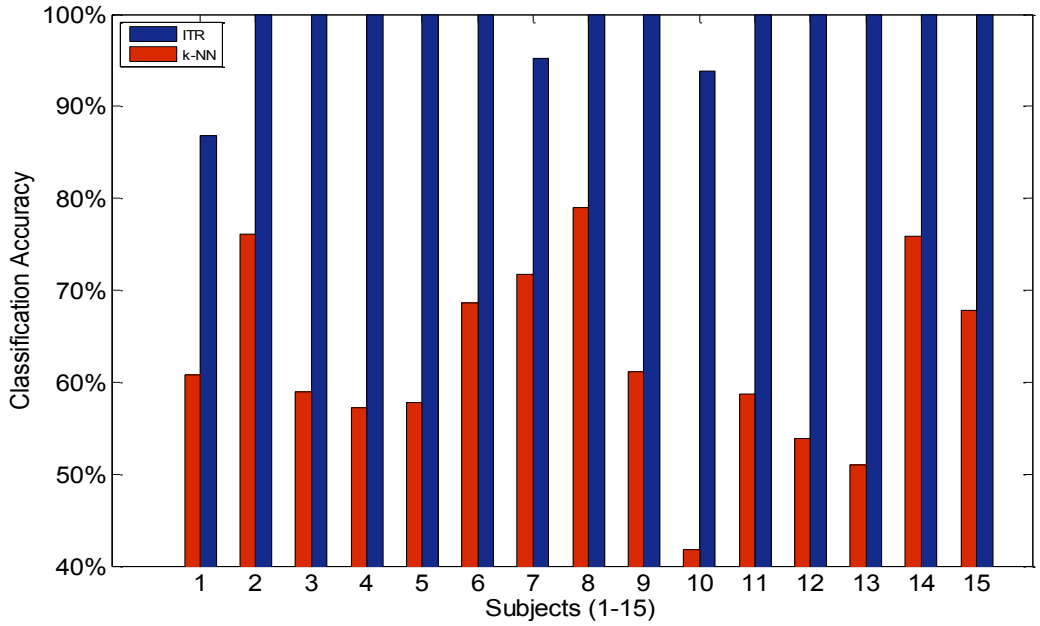


Fig. 8.14 Task classification comparison between ITR and *k*-NN for 15 subjects

It should be noted that the 1-NN in Fig. 8.14 were obtained using the features extracted using a window size of 6400 samples, while the ITR algorithm is still using the original 640 samples per window for feature extraction. Using 6400 samples per window only 12 instances will be generated even after window overlapping (given the 4 minutes of data available), which may reduce the impact of instance selection of the ITR algorithm.

8.4.2 Biometric Performance using MM/I dataset

In this subsection the effectiveness of the proposed algorithm for biometric person recognition using EEG data is investigated and compared with the 1-NN and SVM algorithms. The choice of one important parameter and their impact on performance of the ITR algorithm are explored. The MM/I dataset is used for these investigations of applying for a biometric scenario: data of 105 selected subjects (classes) were employed for classification. All of the biometric tests in this work employed the wavelet-based features [108], which was introduced in Chapter 6.

8.4.2.1 Parameter Optimization

The *Stage 1* of the proposed ITR algorithm involves removing some of the instances from the training set. This investigation is devoted to empirically finding the optimal quantity of the preserved instances for MM/I dataset and the biometrics recognition application. The recognition accuracies were obtained by averaging the results of the leave-one-out three-fold cross-validation (data of 6 minutes of recording in total). Defining the percentage of the retained instances as *P%*, a series of tests were conducted and the recognition results are illustrated in Fig. 8.15.

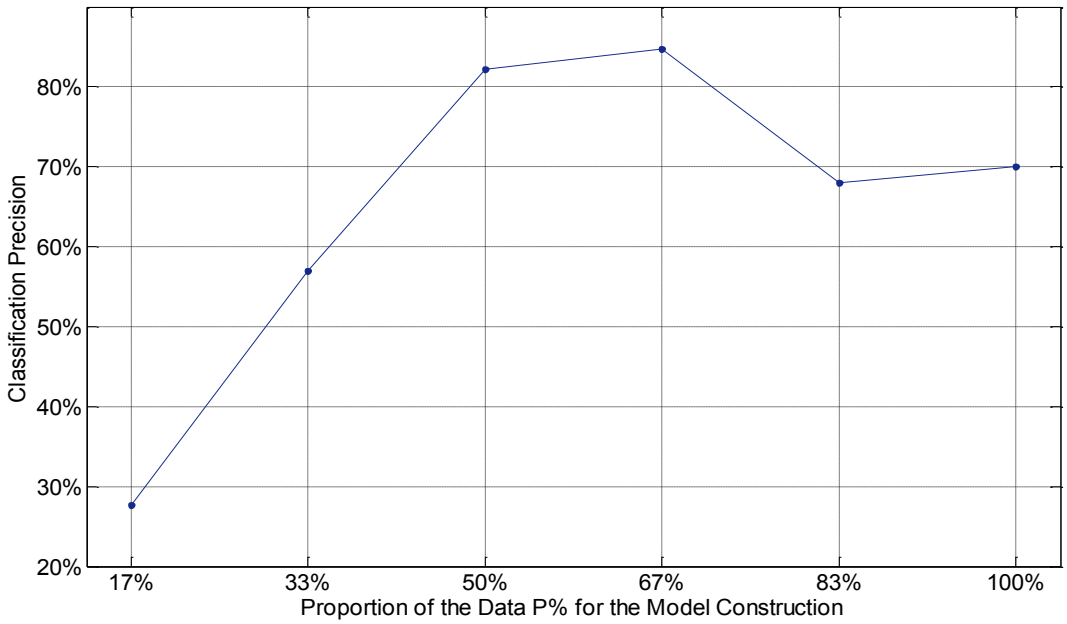


Fig. 8.15 Preliminary tests for optimizing *P%*

The graph in Fig. 8.15 shows the trend of the classification accuracies while *P%* of the available instances was used as the prototype for model construction. As a preliminary test for parameter optimization, only a subset of the MM/I dataset (S1-S10) was used to investigate the general impact of the preserved volume of the training instances. The available instances per class (subject) were grouped by a factor of six (as it is depicted by the horizontal axis in Fig. 8.15) and only the *Stage 1* of the proposed algorithm was applied. Before applying the ITR algorithm, the EEG series was segmented into 240 windows ($\mathrm{I} = 240$), the training set contained two minutes' recording. The preserved data ($\mathrm{J} = 160$) for training are those instances which comprise the elements with the largest inter class distances. The results indicate the best performances tend to be achieved while about half to 2/3 of the available data were kept for training. Therefore, as a rule of thumb about 2/3 of the selected data will be used for the next stage of the algorithm. However, such parameter setting certainly should depend on the quality of the database and the application scenario: for databases containing low quality and noisy signals, such as the Mobile Sensor Database, more instances may be removed in *Stage 1*.

8.4.2.2 Performance Comparison in Identification Scenario

Given the parameter *P%* is empirically optimized (*P%*≈66.67%), the evaluation of the ITR algorithm using the full MM/I dataset is conducted. Data of the selected 105 out of 109 subjects were employed in order to guarantee enough data per subject for processing (three runs, 6 minutes in all). The performance of the two-stage ITR algorithm is compared with *k*-NN (*k*=1) and nonlinear SVM with second order polynomial kernel function [171]. The resulting CMC (cumulative match characteristic [193]) curves are illustrated in Fig. 8.16.

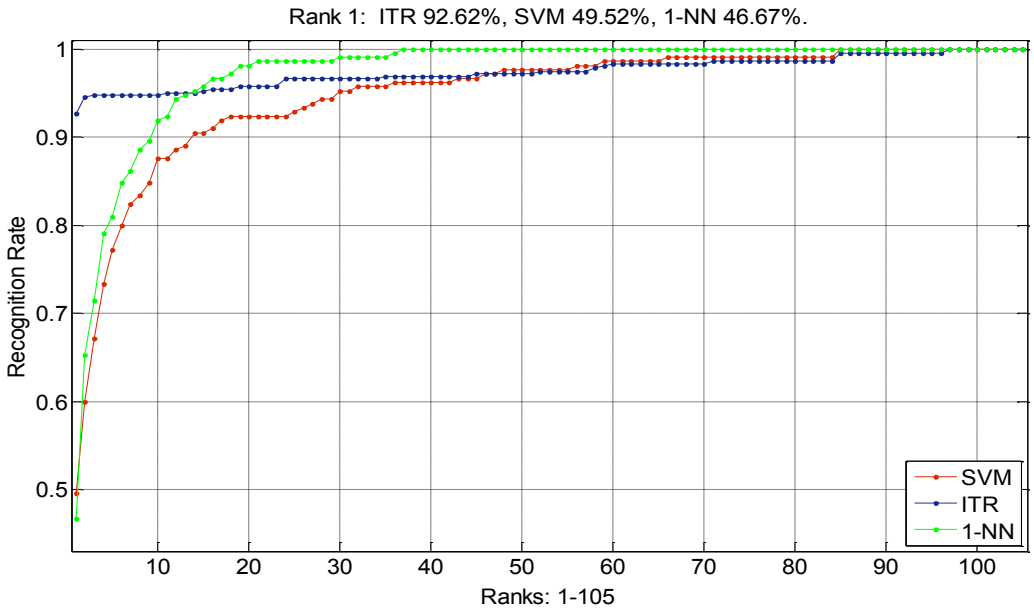


Fig. 8.16 Average recognition performances from different algorithms

The curves in Fig. 8.16 shows ITR algorithm provided the highest rank-1 recognition rate. The 1-NN classifier though provided the lowest rank-1 recognition rate, but its learning speed was found to be the fastest amongst the three algorithms. As expected, the SVM has better rank-1 identification rate than the 1-NN classifier.

As it was mentioned previously, ITR does not only rank and remove the less discriminating/informative instances, but also utilizes all the available elements per class to construct a series of new instances/templates. Therefore, the selected instances are some ranked and re-created new feature vectors, the elements of each vector may stem from different instances of the original dataset. However, the elements of these new templates are still restricted by the dimensions of each instance vector, which guarantees the new templates still possess their original characteristics /meaning.

8.4.2.3 Performance Comparison in Verification Scenario

The effectiveness of the proposed algorithms in the verification scenario was also investigated. The three instance-based learning algorithms ITR, SVM and 1-NN were compared using several DET (detection error trade-off [194]) curves. The results of cross-validation were used to generate the curves. Indeed, from the biometric

verification perspective, the *Stage 1* and *Stage 2* of the proposed algorithm are designed to reduce the FAR and FRR, respectively.

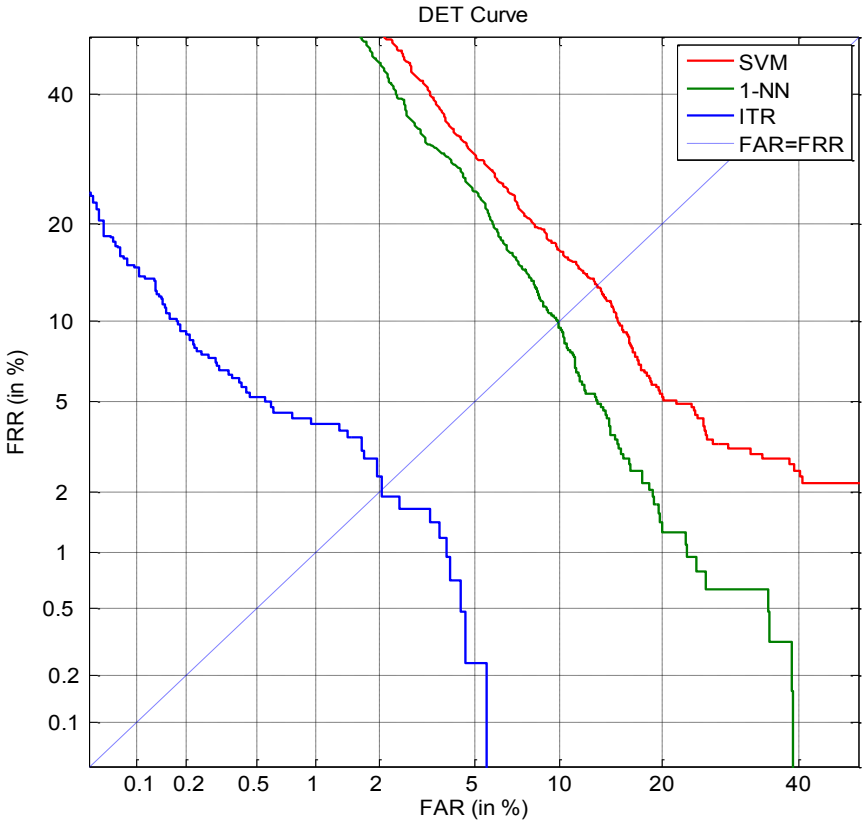


Fig. 8.17 DET curves of three learning algorithms (105 subjects)

The verification performances of the algorithms are illustrated using the DET curves in Fig. 8.17. For the ITR algorithm, each feature vector was generated from an EEG signal recording of 4 seconds (640 samples, at 160 Hz). For the SVM and 1-NN algorithms, this was set as 6400 samples per window.

The graph in Fig. 8.17 shows the equal error rate (EER) for the ITR algorithm as being around 2%; whereas the 1-NN classifier showed an EER of approximately 10%. Surprisingly, SVM classifier provided worse performance than the 1-NN classifier in verification tests, whereas in the identification mode the SVM results better rank-1 accuracy (Fig. 8.15). The results suggest that the ITR algorithm may be particularly suitable for applications where a low FAR is necessary.

8.4.3 Biometric Performance using Mobile Sensor Database

In this section the biometric performance of the proposed algorithm is tested using the self-collected database.

8.4.3.1 Template Ageing Effects

The design of the proposed algorithm was motivated by the need to address the template ageing effect in biometric recognition applications using very limited available EEG data for training and testing. In this section the impact of using EEG data from different sessions, separated in time, using standard classifiers is presented. This is followed by the evaluation of the ITR algorithm in dealing with such data. Finally the performance of the proposed algorithm is compared with the SVM and 1-NN algorithms in both the identification and verification scenarios.

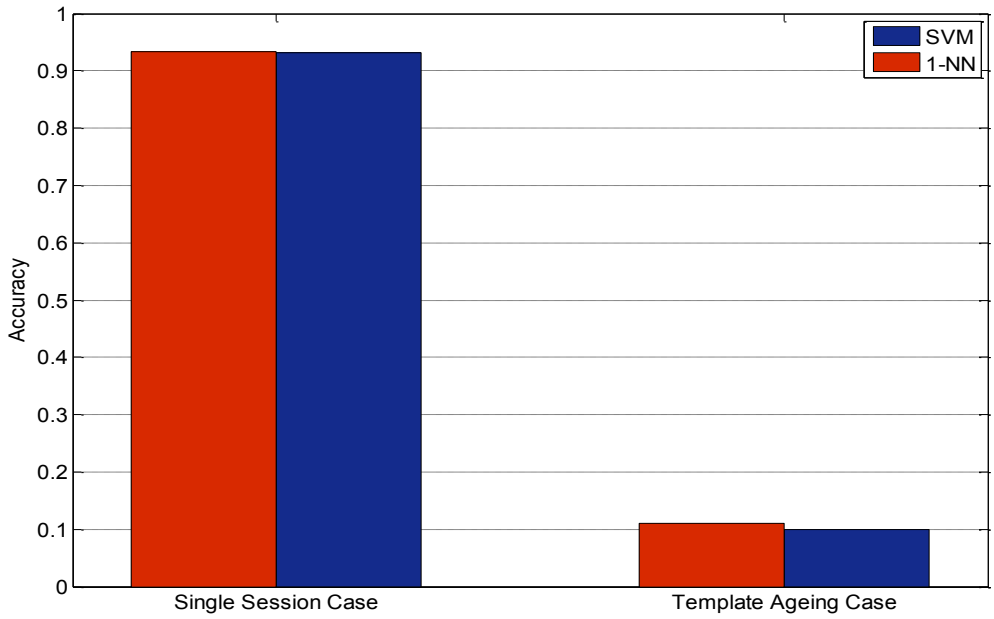


Fig. 8.18 The session impact in EEG biometric identification

Fig. 8.18 illustrates the impact of template ageing when using EEG data for identification. 1-NN and SVM classification with the same parameter settings as the previous section were used for comparison. For the single session case, the one-minute recording was divided into five non-overlapping segments for cross-validation and the average accuracy of the two classification algorithms has been found to be

similar at around 93%. In the Template Ageing Case, data of the entire Session 1 (one minute) was used for training and the data in Session 2 (one minute) used for a five-fold cross-validation test (each segment lasting for 12 seconds). The impact of template ageing is significant as shown to Fig. 8.17: both the SVM and 1-NN provided only about 10% of recognition accuracy. This may be due to the sparse pattern distribution in the feature domain, according to the investigation presented in Section 8.3.

8.4.3.2 Performance Comparison in Identification Scenario

The effectiveness of the proposed algorithm was investigated for the identification scenario; the recognition results obtained using the standard 1-NN and SVM algorithms are presented in Fig. 8.19 for comparison purpose.

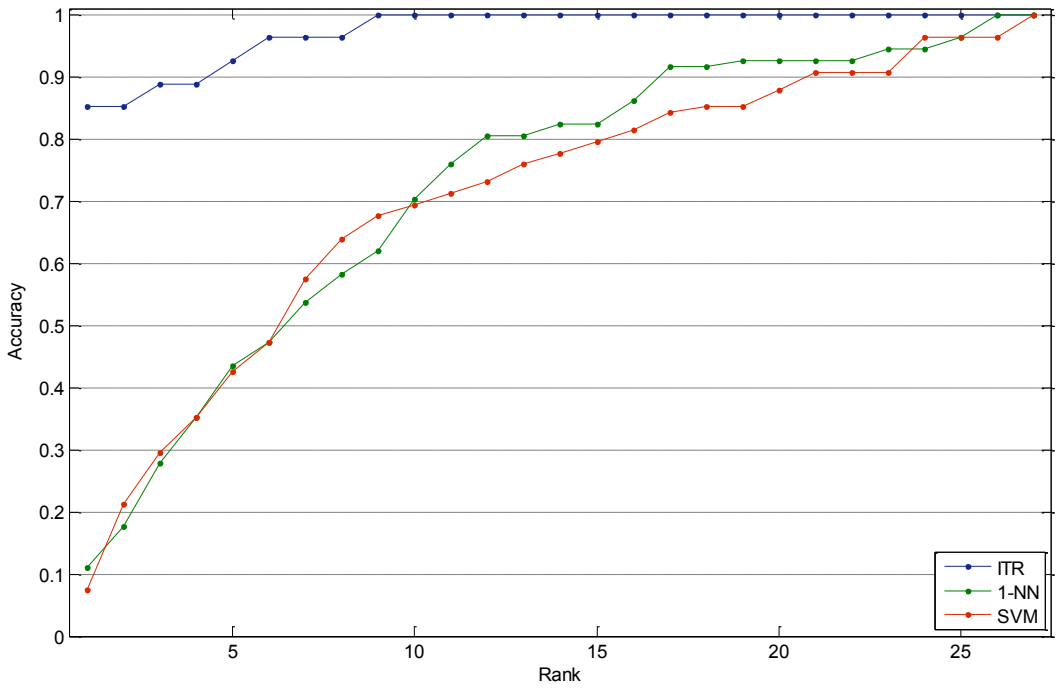


Fig. 8.19 CMC curves of the three learning algorithms

The CMC curves of three classification algorithms were obtained using EEG data of Session 1 (one minute of recording) for model training and the data of 12 seconds' recordings as query sets. The graphs in Fig. 8.19 are obtained using the averaged results tested by five separated recordings of Session 2. The initial parameter $I = 240$

is set for ITR algorithm, after the *Stage 2*, the preserved number of instance is $F = 4$ for both the training set and the query set. As it is depicted in the graph, the ITR method provided much better average rank-1 accuracy (82.1%) than the other classifiers; the SVM and 1-NN provided similar performance.

8.4.3.3 Performance Comparison in Verification Scenario

Fig. 8.20 indicates the DET curves obtained for the verification scenario. Indeed, considering the challenging condition of the experimental design (1 minute of training per class and 12 seconds for testing; using a low-cost single dry sensor system), the EERs of all the algorithms were found to be promising. However, it is clear the ITR algorithm was still able to provide much better performance than the other two classifiers.

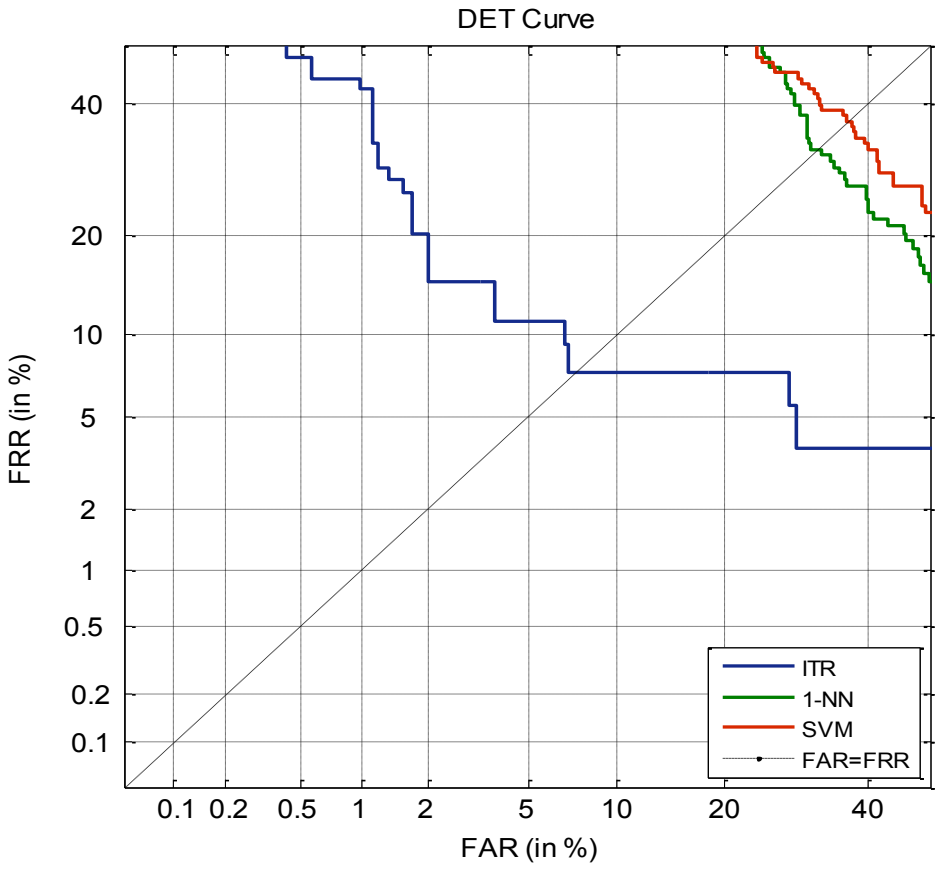


Fig. 8.20 Verification performance of the ITR algorithm

According to the graphs in Fig. 8.19, the EER of ITR method is found to be around 7.5%; whereas both of the other two algorithms had substantially higher EERs. The results in Fig. 8.17 and Fig. 8.20 indicate the difference of the MM/I dataset and the Mobile Sensor Database in terms of biometric data quality: even by applying the ITR algorithm the lowest EER for the MM/I (105 subjects) was able to reach round 2% whereas for the Mobile Sensor Database (27 subjects) it was around 7.5%.

The effectiveness of the ITR algorithm in removing the potentially less informative instances is also considerable: for MM/I dataset the EERs difference between ITR and 1-NN is about 5% (2% versus 10%); for Mobile Sensor Database such difference is about 25% (7.5% versus 33%), which seems to have verified the low-cost sensor system in less-controlled environment captured much more noisy EEG signals than BCI2000.

8.5 Discussion and Conclusion

In this section some remaining issues of the ITR algorithm will be further clarified. This will be followed by overall conclusions and suggestions for further work.

8.5.1 Discussion

Based on the rationale presented for the ITR algorithm, the available feature vectors have been decomposed and the elements reconstructed into new instance vectors. It may be argued that such operation (distort the time order of the vector elements) could potentially destroy the biometric specificity of the feature vectors (multiple dimensions). However, as it was clarified in Section 8.2, the ITR does not create new measurements, only observing the existing measurements (both training set and query) in the further detailed level: instead of handling instance in vector level, it rearranges with the elements of the instance vectors.

The performance is also governed by the employed features. To highlight the important role of the feature played in ITR-based classification, several tests were further conducted using different features to investigate the above-mentioned arguments from the experimental perspective. Multiple wavelet-based features are

used to test the effectiveness of different features using the ITR algorithm. The motivation of these tests is to justify the following statement: 1) biometric characteristics of the data are preserved by the features extracted for feature extraction; depending on the type of the features the recognition performance varies a lot. 2) The locations of the elements while constructing new feature vectors is not necessarily kept the same as the original (time domain order). It should be mentioned that the window size during the feature extraction also plays an important role: the feature computed based relatively bigger window sizes generally tend to preserve more biometric characteristics. In the extreme case, were the ITR algorithm applied to each sample (window size: one sample/window), the algorithm may destroy the time-domain characteristic of the data but may not necessarily lead to bad biometric performance.

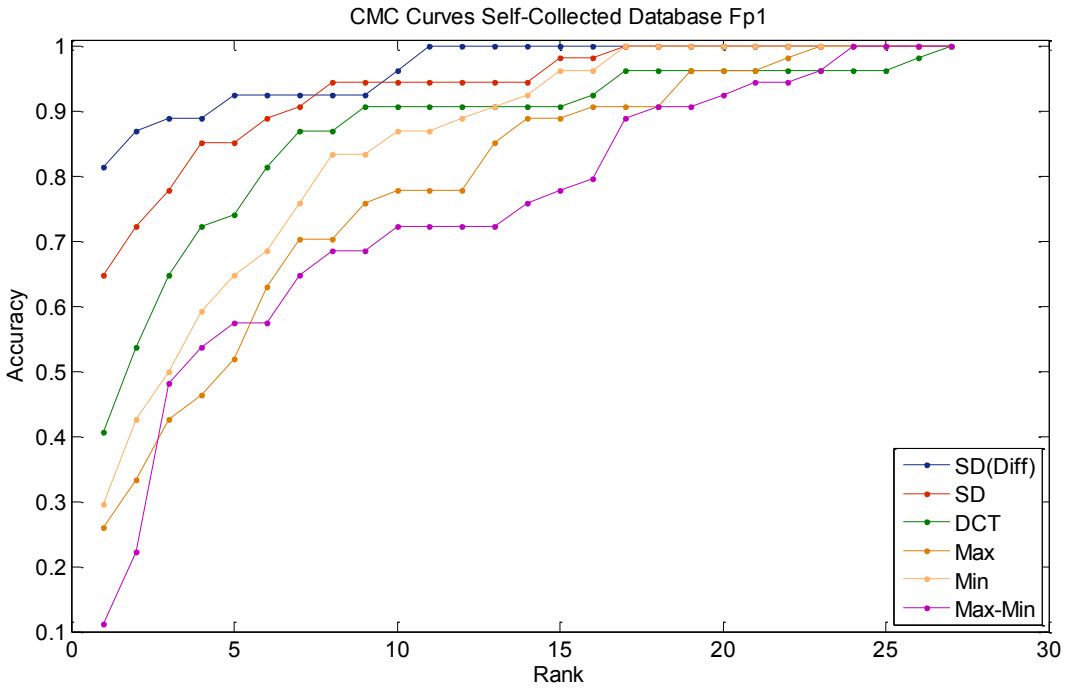


Fig. 8.21 Effectiveness of different features using ITR algorithm

Fig. 8.21 shows the CMC curves of different wavelet-based features obtained using the ITR algorithm, tested with the Mobile Sensor Database. The window size is kept the same as 512 samples/ window (1 second). If the measurements lost their specificities by applying ITR algorithm, different types of feature vectors would have

provided similar result. Considering the obtained different ranks (from rank 1 to rank 27) accuracies using different features in Fig. 8.20, it is clear that the type of the extracted feature plays an important role, which indicates the ITR algorithm preserves the characteristics of the original measurements captured by different types of features, in the same way as other conventional instance reduction/learning algorithms. These results indicate the EEG-based biometric recognition is not necessarily depending on the order of the features (location of the windows in the time domain).

Indeed, the ITR algorithm's rationale dictates that this method may not be quite suitable for applying directly to the original raw signal for classification: depending on the selected window size it may be unable to capture enough information of the signal if applied directly (for example, 1 sample/window). However, it is related to window size and feature rather than feature modelling. Therefore, to achieve good classification performance, windowing, feature extraction and model construction are equally important in EEG-based experimental design.

Fig. 8.21 shows the CMC curves of different wavelet-based features obtained using the ITR algorithm, tested with the Mobile Sensor Database. The window size is kept the same as 512 samples/ window (1 second). If the measurements lost their specificities by applying ITR algorithm, different types of feature vectors would have provided similar and optimised result. Considering the obtained different rank-1 accuracies and regression speeds using different features in Fig 8.20, it is clear the type of the extracted feature plays an extremely important role, which indicates the ITR algorithm preserves the characteristics of the original measurements captured by different types of features, in the same way as other conventional instance reduction/learning algorithms. These results indicate the EEG-based biometric recognition is not necessarily depending on the order of the features (order of the windows in the time domain); feature extraction plays an important role in the biometric recognition.

Indeed, the ITR algorithm's rationale dictates that this method may not be quite suitable for applying directly to the original raw signal for classification: depend on

the selected window size it may be unable to capture enough biometric information of the signal if applied directly (for example, 1 sample/window). However, such incompetence precisely is related to window size and feature rather than feature modelling. Therefore, to achieve good classification performance, windowing, feature extraction and model construction are equally important in EEG-based experimental design.

8.5.2 Conclusions

In this chapter a new instance-based template reconstruction learning algorithm was presented and its rationale explained. This was followed by three case studies evaluating the effectiveness of the algorithm. The performance of the ITR algorithm was compared with two other popular instance-based algorithms namely *k*-NN and SVM using two EEG databases. The proposed algorithm has been shown to result in significant performance improvements in different scenarios. The proposed algorithm is especially good at dealing with the template ageing effects of non-stationary time series with small amounts of available data, owing to its effectiveness of adaptive learning in *Stage 2*. The ITR algorithm also can be viewed as a black box to increase the regularity of the existing features: as with the reducing of the window size for feature extraction, the entropy of the resulting features will increase. Therefore, the proposed ITR algorithm is a generic framework: the purpose of template reconstruction for better recognition could also be achieved by other methods. For example, the Bayesian methods may be used to optimize the feature vectors as well. The effectiveness of the proposed algorithm is expected to be tested using other databases to further prove its generality. The next chapter is devoted to the conclusions of the thesis and some future work for EEG biometrics.

Chapter 9

Conclusions and Future Work

A summary of the work reported in the thesis is presented in this chapter. Major contributions to the research are described first. This is then followed by a discussion and suggestion for further work regarding using EEG for biometric recognition.

9.1 Contribution

This section is devoted to highlighting the contributions of this thesis. The major contributions are related to their respective chapters.

In short the contributions of this work include:

- a new metric for the usability of EEG based biometric systems
- a new database of EEG data obtained using low-cost mobile sensor with the potential for investigating template ageing effects
- an investigation into task sensitivity of EEG data for biometric applications
- a new pre-processing scheme for enhanced noise removal
- novel features for extracting biometric information from EEG data
- a novel feature filtering technique
- a novel feature classification algorithm

9.2 Discussions and Conclusions

As a biometric modality, the implementation of EEG signals affected by multiple factors, metrics such as the accuracy rates (*CRR*, *FAR* and *FRR*) alone may not be enough to represent its performance. Four other factors were considered in the thesis to evaluate the performance of a biometric system: 1) the involved number of the subjects; 2) number of electrodes employed; 3) the size of the training set and 4) the size of the test set. Along with the accuracy, a performance score was computed using

all the five factors. However, the proposed metric does have its liability: the influential template ageing/changing effects is not represented.

To address the problem of template ageing in EEG biometrics, a new database was collected in addition to facilitating two publicly available databases. The new database (Mobile Sensor Database) contains the data of 27 subjects in two separated recording sessions (per subject), with the time interval between the sessions ranging from three weeks to two months and the data of 50 subjects with only single session. Mobile Sensor Database was collected using a low-cost headset with a single dry sensor. The data of each session lasted only one minute with subjects mentally performing a simple number counting task. This database also helps to investigate the feasibility of using EEG signals in a realistic scenario with an easy-to-deploy sensor.

Many scattered but important issues regard using EEG signals for biometric recognition were investigated experimentally. It worth to mention that, though conventionally the informative frequency range of EEG signals is between 0 to 50 Hz or 60 Hz, in its biometrics implementation, however, it is found in this investigation that higher part of the frequency (60 Hz-80 Hz) also effective [7].

The impact of different type of EEG signals in biometric recognition was investigated, which potentially may alleviate the template ageing effects. A training data accumulation scheme for model construction was developed which combines data from different types of motor movement/imagery tasks (from separated recordings) to produce more effective training using the available data. The results indicate when using separated recordings for training and test respectively, the once considerable preformation degradation is non-effective by employing the proposed scheme.

Four new EEG-based biometric features were proposed in the thesis. Two of them are based on HHT and the other two are wavelet-based features. The features derived from HHT namely InsAmp and InsFreq features provided the best performance (InsAmp producing the highest performance above all); the wavelet features are 1) the derivative of wavelet coefficients and 2) wavelet-DCT coefficient feature. The computation for wavelet features is much faster than HHT features, however, wavelet features achieved a lower accuracy in both identification and verification scenarios.

The proposed wavelet-based features performed better than some other conventional features for the related databases.

The HHT-based features were derived from Hilbert Transform and Empirical Mode Decomposition, which is designed for better handling of non-stationary data (such as EEG signals). The introduction of the differentiation stage in the processing of wavelet-based coefficients is motivated by the role played by differentiation in some image processing tasks (e.g. edge enhancement), which indicates the directional change of the intensity or colour in an image. The Wavelet-DCT feature, on the other hand, is motivated by one standard algorithm in voice recognition field which uses Mel-Frequency Cepstral Coefficients (MFCC) as feature for recognition. Of the proposed two wavelet-based features, the coefficients differentiation feature provided better Rank 1 recognition accuracy and the wavelet-DCT feature showed a faster regression speed in the CMC curve.

A novel quality filtering algorithm was proposed, which employs Sample Entropy (SampEn) as an indicator to predict and preserve the high quality segments/windows of each class for feature extraction. As a feature filtering method, the recognition accuracy improved by 4%~5% after filtering and only about 10% of all the templates were preserved for model construction.

An Instance-based Template Reconstruction (ITR) learning algorithm was proposed, which is designed to resolve both the inter-class and intra-class problems in pattern recognition. As a generic algorithm, the proposed two-stage ITR algorithm is designed to optimise the training model by better separating the between-class patterns (*Stage 1*) and, adaptively enhance the within-class similarity between the training set and test set (*Stage 2*).

The ITR algorithm was designed to alleviate the template ageing effects while using EEG signals for biometric recognition. In the meantime, the adaptive template reconstruction mechanism also allows a significant volume reduction for both training and test sets. Additionally, the proposed algorithm is relatively insensitive to the quality of the available data. Therefore, it may be especially good at processing

low quality time-series data (such as the EEG signals collected through low-cost sensors).

9.3 Further Work

As it was reported in the Feature Extraction chapter (Chapter 7), the HHT-based features provided better performance than wavelet-based features but much less efficient in computing. One future work will be optimising the HHT-based algorithm in programming, since the current code is designed for algorithm implementation without considering the efficiency of computing. It is hoped the computation of EMD-based algorithms could be comparable with wavelet algorithms.

The Mobile Sensor Database contains 27 subjects with double sessions. It is interesting to investigate the effectiveness of the proposed algorithms (especially the one in Chapter 5 and the one in Chapter 8) on its increased version in the future: 1) increase the number of sessions and 2) increase the number of subjects.

The stimuli of triggering the EEG signals for biometric recognition can be further investigated. The audial stimulus is such a potential candidate: by far there has been only one report regarding imaging audial sounds for EEG biometric recognition, which is actually not real audial stimulus.

REFERENCES

[1] A. Jain, L. Hong, and S. Pankanti, "Biometric identification," *Commun. ACM*, vol. 43, no. 2, pp. 90–98, Feb. 2000.

[2] A. K. Jain, P. Flynn, and A. A. Ross, Eds., *Handbook of Biometrics*. Boston, MA: Springer US, 2008.

[3] L. F. Haas, "Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography," *J. Neurol. Neurosurg. Psychiatry*, vol. 74, no. 1, pp. 9–9, Jan. 2003.

[4] W. H. Miltner, C. Braun, M. Arnold, H. Witte, and E. Taub, "Coherence of gamma-band EEG activity as a basis for associative learning.," *Nature*, vol. 397, no. 6718, pp. 434–6, Feb. 1999.

[5] M. Poulos, M. Rangoussi, V. Chrissikopoulos, and A. Evangelou, "Parametric person identification from the EEG using computational geometry," in *ICECS'99. Proceedings of ICECS '99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.99EX357)*, 1999, vol. 2, no. 2, pp. 1005–1008.

[6] M. Poulos, M. Rangoussi, and N. Alexandris, "Neural network based person identification using EEG features," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, 1999, vol. 2, pp. 1117–1120 vol.2.

[7] S. Yang and F. Deravi, "Novel HHT-Based Features for Biometric Identification Using EEG Signals," *2014 22nd Int. Conf. Pattern Recognit.*, pp. 1922–1927, Aug. 2014.

[8] *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2005.

[9] B. E. Swartz, "The advantages of digital over analog recording techniques," *Electroencephalogr. Clin. Neurophysiol.*, vol. 106, no. 2, pp. 113–117, Feb. 1998.

[10] A. Coenen, E. Fine, and O. Zayachkivska, "Adolf Beck: a forgotten pioneer in electroencephalography.," *J. Hist. Neurosci.*, vol. 23, no. 3, pp. 276–86, Jan. 2014.

[11] S. Aydin, H. M. Saraoğlu, and S. Kara, "Log energy entropy-based EEG classification with multilayer neural networks in seizure.," *Ann. Biomed. Eng.*, vol. 37, no. 12, pp. 2626–30, Dec. 2009.

[12] P. Campisi and D. La Rocca, "Brain waves for automatic biometric based user recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 6013, no. c, pp. 1–1, 2014.

[13] S. N. Abbas, M. Abo-Zahhad, and S. M. Ahmed, "State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals," *IET Biometrics*, Mar. 2015.

[14] R. W. Homan, J. Herman, and P. Purdy, “Cerebral location of international 10–20 system electrode placement,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 4, pp. 376–382, Apr. 1987.

[15] “EEG Motor Movement/Imagery Dataset.” [Online]. Available: http://www.physionet.org/pn4/eegmmidb/. [Accessed: 20-Apr-2015].

[16] G. CHATRIAN, “Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity,” *Am J Electroencephalogr Technol*, vol. 25, pp. 83–92, 1985.

[17] “Neurosky Products.” [Online]. Available: http://store.neurosky.com/products. [Accessed: 20-Apr-2015].

[18] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, “BCI2000: a general-purpose brain-computer interface (BCI) system.,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–43, Jun. 2004.

[19] schalklab, “BCI 2000.” [Online]. Available: http://www.schalklab.org/research/bci2000. [Accessed: 21-May-2015].

[20] “Emotiv.” [Online]. Available: http://emotiv.com/. [Accessed: 20-Apr-2015].

[21] “MindWave.” [Online]. Available: http://store.neurosky.com/products/mindwave-1. [Accessed: 10-Jul-2015].

[22] “OCZ NIA Neural Impulse Actuator: Amazon.co.uk: Computers & Accessories.” [Online]. Available: http://www.amazon.co.uk/OCZ-NIA-Neural-Impulse-Actuator/dp/B00168VU4U. [Accessed: 10-Jul-2015].

[23] “Melon Headband.” [Online]. Available: http://www.thinkmelon.com/. [Accessed: 10-Jul-2015].

[24] “HiBrain.” [Online]. Available: http://www.hyperneuro.cn/. [Accessed: 10-Jul-2015].

[25] “FocusBand - Mind Training Headset.” [Online]. Available: http://www.ifocusband.com/. [Accessed: 10-Jul-2015].

[26] “MUSE ™ | Meditation Made Easy.” [Online]. Available: http://www.choosemuse.com/. [Accessed: 10-Jul-2015].

[27] “OpenBCI.” [Online]. Available: http://www.openbci.com/. [Accessed: 10-Jul-2015].

[28] “Aurora Dream Headband.” [Online]. Available: https://iwinks.org/. [Accessed: 10-Jul-2015].

[29] W. O. Tatum, “Ellen R. Grass Lecture: Extraordinary EEG,” *Neurodiagn. J.*, Jan. 2014.

[30] F. Su, L. Xia, A. Cai, Y. Wu, and J. Ma, “EEG-based Personal Identification: from Proof-of-Concept to A Practical System,” in *2010 20th International Conference on*

Pattern Recognition, 2010, pp. 3728–3731.

[31] D. LA Rocca, P. Campisi, B. Vegso, P. Cserti, G. Kozmann, F. Babiloni, and F. DE Vico Fallani, "Human brain distinctiveness based on EEG spectral coherence connectivity.," *IEEE Trans. Biomed. Eng.*, vol. 9294, no. c, pp. 1–7, Apr. 2014.

[32] M. Fraschini, A. Hillebrand, M. Demuru, L. Didaci, and G. L. Marcialis, "An EEG-Based Biometric System Using Eigenvector Centrality in Resting State Brain Networks," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 666–670, Jun. 2015.

[33] S. J. Luck, *An Introduction to the Event-Related Potential Technique*. MIT Press, 2014.

[34] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 174–179, Jun. 2000.

[35] J. Polich, "Updating P300: an integrative theory of P3a and P3b.," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–48, Oct. 2007.

[36] R. Palaniappan and P. Raveendran, "Individual identification technique using visual evoked potential signals," *Electron. Lett.*, vol. 38, no. 25, p. 1634, Dec. 2002.

[37] J. G. Snodgrass and M. Vanderwart, "A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity.," *J. Exp. Psychol. Hum. Learn. Mem.*, vol. 6, no. 2, pp. 174–215, 1980.

[38] R. Palaniappan, "A New Method to Identify Individuals Using Signals from the Brain," in *Proceedings of the Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, and Fourth Pacific Rim Conference on Multimedia*, 2003, pp. 1442–1445.

[39] R. Palaniappan, "Recognising Individuals Using Their Brain Patterns," in *Third International Conference on Information Technology and Applications (ICITA'05)*, 2005, vol. 2, pp. 520–523.

[40] Palaniappan, "Vision Related Brain Activity for Biometric Authentication," 2006, pp. 3227–3231.

[41] R. Palaniappan and D. P. Mandic, "EEG Based Biometric Framework for Automatic Identity Verification," *J. VLSI Signal Process. Syst. Signal Image. Video Technol.*, vol. 49, no. 2, pp. 243–250, Jun. 2007.

[42] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: a machine learning approach.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 738–42, Apr. 2007.

[43] R. Palaniappan, "Identifying Individuality Using Mental Task Based Brain Computer Interface," in *Third International Conference on Intelligent Sensing and Information Processing, ICISIP*, 2005, pp. 238–242.

[44] S. Marcel and J. D. R. Millán, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 743–52, Apr. 2007.

[45] Q. Gui, Z. Jin, M. V. R. Blondet, and S. Laszlo, "Towards EEG Biometrics : Pattern Matching Approaches for User Identification," in *2015 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2015, pp. 1–6.

[46] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I Think , Therefore I Am : Usability and Security of Authentication Using Brainwaves," vol. 0424422, pp. 1–16, 2013.

[47] A. Oppenheim, *Discrete-time signal processing*. Englewood Cliffs N.J.: Prentice Hall, 1989.

[48] R. Palaniappan, "Individual identification technique using visual evoked potential signals," *Electron. Lett.*, vol. 38, no. 25, pp. 1634–1635, 2002.

[49] S. Marcel and J. D. R. Millán, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 743–52, Apr. 2007.

[50] G. Safont, A. Salazar, A. Soriano, and L. Vergara, "Combination of multiple detectors for EEG based biometric identification/authentication," in *2012 IEEE International Carnahan Conference on Security Technology (ICCST)*, 2012, pp. 230–236.

[51] C. Miyamoto, S. Baba, and I. Nakanishi, "Biometric person authentication using new spectral features of electroencephalogram (EEG)," in *2008 International Symposium on Intelligent Signal Processing and Communications Systems*, 2009, pp. 1–4.

[52] I. Nakanishi, S. Baba, C. Miyamoto, and A. B. Wave, "EEG Based Biometric Authentication Using New Spectral Features," 2009, no. Ispacs, pp. 651–654.

[53] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, 1974.

[54] R. B. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles', "The electroencephalogram as a biometric," in *Canadian Conference on Electrical and Computer Engineering 2001. Conference Proceedings (Cat. No.01TH8555)*, 2001, vol. 2, pp. 1363–1366.

[55] A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau, and G. Ruffini, "Unobtrusive Biometric System Based on Electroencephalogram Analysis," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, p. 143728, 2008.

[56] K. Brigham and B. V. K. V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2010, pp. 1–8.

[57] P. Campisi, G. Scarano, F. Babiloni, F. D. V. Fallani, S. Colonnese, E. Maiorana, L.

Forastiere, L. Sapienza, U. Roma, and V. Eudossiana, “Brain waves based user recognition using the ‘ Eyes Closed Resting Conditions ’ protocol,” 2011, vol. 00, no. c, pp. 16–19.

[58] Z. Dan, Z. Xifeng, and G. Qiangang, “An Identification System Based on Portable EEG Acquisition Equipment,” in *2013 Third International Conference on Intelligent System Design and Engineering Applications*, 2013, pp. 281–284.

[59] I. Daubechies, *Ten lectures on wavelets*, Vol. 61. Philadelphia: Society for industrial and applied mathematics, 1992.

[60] C. N. Gupta, Y. U. Khan, R. Palaniappan, and F. Sepulveda, “Wavelet Framework for Improved Target Detection in Oddball Paradigms Using P300 and Gamma Band Analysis,” vol. 14, no. 2, pp. 61–67, 2009.

[61] M. K. Abdullah, K. S. Subari, J. Leo, C. Loong, and N. N. Ahmad, “Analysis of the EEG Signal for a Practical Biometric System,” in *World Academy of Science, Engineering and Technology*, 2010, no. 2008, pp. 1123–1127.

[62] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[63] P. Flandrin, G. Rilling, and P. Gonçalvés, “Empirical Mode Decomposition as a Filter Bank,” *Signal Process. Lett.*, vol. 11, no. 2, pp. 112–114, 2004.

[64] P. Kumarii, S. Kumar, and A. Vaishi, “Feature Extraction using Emprical Mode Decomposition for Biometric System,” in *Signal Propagation and Computer Technology (ICSPCT)*, 2014, pp. 283–287.

[65] G. K. Singhal and P. RamKumar, “Person Identification Using Evoked Potentials and Peak Matching,” in *2007 Biometrics Symposium*, 2007, pp. 1–6.

[66] X. Huang, “Human Identification with Electroencephalogram (EEG) Signal Processing,” in *International Symposium on Communications and Information Technologies (ISCIT)*, 2012, pp. 1021–1026.

[67] S.-W. L. Seul-Ki Yearn, Heung-II Suk, “EEG-based Person Authentication using Face Stimuli,” in *International Winter Workshop on Brain-Computer Interface (BCI)*, 2013, pp. 58–61.

[68] D. Phung, D. Tran, W. Ma, P. Nguyen, and T. Pham, “Using Shannon Entropy as EEG Signal Feature for Fast Person Identification,” in *European Symposium on Artificial Neural Networks (ESANN)*, 2014, pp. 23–25.

[69] P. Nguyen, D. Tran, X. Huang, and D. Sharma, “A Proposed Feature Extraction Method for EEG-based Person Identification,” in *International Conference on Artificial Intelligence (ICAI)*, 2012.

[70] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, “Machine learning: A review of

classification and combining techniques," *Artif. Intell. Rev.*, vol. 26, pp. 159–190, 2006.

[71] S. J. Russell, P. Norvig, J. F. Canny, J. M. Malik, D. D. Edwards, and S. J. S. Jonathan, *Artificial Intelligence: A Modern Approach*, vol. 13, no. 2. Prentice-Hall, Egnlewood Cliffs 25, 1995.

[72] F. Nigsch, A. Bender, B. van Buuren, J. Tissen, E. Nigsch, and J. B. O. Mitchell, "Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization.," *J. Chem. Inf. Model.*, vol. 46, no. 6, pp. 2412–22, Jan. 2006.

[73] R. Palaniappan and K. V. R. Ravi, "Improving visual evoked potential feature classification for person recognition using PCA and normalization," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 726–733, May 2006.

[74] A. Yazdani, A. Roodaki, S. H. Rezatofighi, K. Misaghian, and S. K. Setarehdan, "Fisher linear discriminant based person identification using visual evoked potentials," in *2008 9th International Conference on Signal Processing*, 2008, pp. 1677–1680.

[75] F. Su, L. Xia, A. Cai, and J. Ma, "Evaluation of recording factors in EEG-based personal identification: A vital step in real implementations," in *2010 IEEE International Conference on Systems, Man and Cybernetics*, 2010, pp. 3861–3866.

[76] G. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley & Sons, 2004.

[77] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, Sep. 1936.

[78] R. Palaniappan, "Electroencephalogram Signals from Imagined Activities : A Novel Biometric Identifier for a Small Population," *Intelligent Data Engineering and Automated Learning–IDEAL 2006*, pp. 604–611, 2006.

[79] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, 2001.

[80] H. J. Lee, H. S. Kim, and K. S. Park, "A study on the reproducibility of biometric authentication based on electroencephalogram (EEG)," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2013, pp. 13–16.

[81] J. S. Kostilek, M., "EEG biometric identification : repeatability and influence of movement-related EEG," in *International Conference on Applied Electronics (AE)*, 2012, pp. 147–150.

[82] and D. G. S. R.O. Duda, P.E. Hart, "Pattern Classification," *J. Classif.*, vol. 24, no. 2, pp. 305–307, Sep. 2007.

[83] R. Palaniappan, "Method of identifying individuals using VEP signals and neural network," in *Science, Measurement and Technology*, 2004, vol. 151, no. 1, pp. 16–20.

[84] R. Palaniappan and D. P. Mandic, “Energy of Brain Potentials Evoked During Visual Stimulus : A New Biometric ?,” in *Artificial Neural Networks: Formal Models and Their Applications–ICANN*, 2005, pp. 735–740.

[85] F. Chunying, L. Haifeng, M. Lin, and J. Bing, “Induced Event-Related Coherence Measures during Auditory Change Detection,” in *2014 International Conference on Medical Biometrics*, 2014, pp. 118–124.

[86] R. Palaniappan, J. Gosalia, K. Revett, and A. Samraj, “PIN Generation Using Single Channel EEG Biometric,” in *Advances in Computing and Communications*, 2011, pp. 378–385.

[87] Q. Gui and Z. Jin, “Exploring EEG-based Biometrics for User Identification and Authentication,” in *Signal Processing in Medicine and Biology Symposium (SPMB)*, 2014, pp. 1–6.

[88] F. Minow, “EASY CAP | EEG Recording Caps and Related Products | FMS | Falk Minow Services.” FMS - Falk Minow Services - www.easycap.de.

[89] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, “A novel connectionist system for unconstrained handwriting recognition.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–68, May 2009.

[90] D. Ciresan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642–3649.

[91] A. Aizerman, E. Braverman, and L. Rozoner, “Theoretical foundations of the potential function method in pattern recognition learning,” *Autom. Remote Control*, vol. 25, pp. 821 – 837, 1964.

[92] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[93] N. C. Oza, R. Polikar, J. Kittler, and F. Roli, Eds., *Multiple Classifier Systems*, vol. 3541. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[94] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines.,” *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–25, Jan. 2002.

[95] H. U. Jian-feng, “Comparison of Different Classifiers for Biometric System Based on EEG signals,” 2010, no. 2, pp. 288–291.

[96] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein, “Low-cost electroencephalogram (EEG) based authentication,” in *2011 5th International IEEE/EMBS Conference on Neural Engineering*, 2011, pp. 442–445.

[97] S. Sun, “Multitask learning for EEG-based biometrics,” in *2008 19th International Conference on Pattern Recognition*, 2008, pp. 1–4.

[98] C. R. Hema and a a Osman, “Single trial analysis on EEG signatures to identify

individuals," in *2010 6th International Colloquium on Signal Processing & its Applications*, 2010, pp. 1–3.

[99] Q. Zhao, H. Peng, B. Hu, Q. Liu, and L. Liu, "Improving Individual Identification in Security Check with an EEG Based Biometric Solution," *Brain Informatics, Springer Berlin Heidelb.*, pp. 145–155, 2010.

[100] B. Quintela and S. Cunha, "Biometric Authentication Using Brain Responses to Visual Stimuli," in *In BIOSIGNALS*, 2010, pp. 103–112.

[101] Y. Bai, Z. Zhang, and D. Ming, "Feature selection and channel optimization for biometric identification based on visual evoked potentials," in *19th International Conference on Digital Signal Processing (DSP)*, 2014, no. August, pp. 772–776.

[102] M. Poulos, M. Rangoussi, V. Chrissikopoulos, and A. Evangelou, "Parametric person identification from the EEG using computational geometry," in *ICECS'99. Proceedings of ICECS '99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.99EX357)*, 1999, vol. 2, pp. 1005–1008.

[103] M. Poulos, M. Rangoussi, N. Alexandris, and a Evangelou, "Person identification from the EEG using nonlinear signal classification.," *Methods Inf. Med.*, vol. 41, no. 1, pp. 64–75, Jan. 2002.

[104] D. La Rocca, P. Campisi, and G. Scarano, "EEG Biometrics for Individual Recognition in Resting State with Closed Eyes," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2012, pp. 1–12.

[105] S. Yang and F. Deravi, "On the Effectiveness of EEG Signals as a Source of Biometric Information," in *2012 Third International Conference on Emerging Security Technologies*, 2012, pp. 49–52.

[106] S. Yang and F. Deravi, "Quality Filtering of EEG Signals for Enhanced Biometric Recognition," in *International Conference on Biometrics Special Interest Group (BIOSIG)*, 2013.

[107] D. La Rocca, P. Campisi, and J. Sol, "EEG Based User Recognition Using BUMP Modelling," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2013, pp. 1–12.

[108] S. Yang and F. Deravi, "Wavelet-Based EEG Preprocessing for Biometric Applications," *2013 Fourth Int. Conf. Emerg. Secur. Technol.*, pp. 43–46, Sep. 2013.

[109] M. V. R. Blondet and S. Laszlo, "Assessment of Permanence of Non-volitional EEG Brainwaves as a Biometric," in *2015 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2015, pp. 1–6.

[110] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.

[111] "UCI Machine Learning Repository: EEG Database Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/EEG+Database. [Accessed: 22-Apr-2015].

[112] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[113] D. M. Powers, "Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation." Bioinfo Publications, 15-Dec-2011.

[114] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," *Perform. Eval. Track. Surveill. (PETS), 10th Int. Work.*, vol. 3, pp. 41–47, 2007.

[115] G. O. Williams, "Iris recognition technology," in *Proceedings of IEEE International Carnahan Conference on Security Technology*, pp. 46–59.

[116] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," 1995, pp. 1137–1143.

[117] S. Marella, "EEG artifacts." [Online]. Available: http://www.slideshare.net/SudhakarMarella/eeg-artifacts-15175461. [Accessed: 21-Apr-2015].

[118] E. Kirmizi-Alsan, Z. Bayraktaroglu, H. Gurvit, Y. H. Keskin, M. Emre, and T. Demiralp, "Comparative analysis of event-related potentials during Go/NoGo and CPT: decomposition of electrophysiological markers of response inhibition and sustained attention.," *Brain Res.*, vol. 1104, no. 1, pp. 114–28, Aug. 2006.

[119] M. A. Kisley and Z. M. Cornwell, "Gamma and beta neural activity evoked during a sensory gating paradigm: effects of auditory, somatosensory and cross-modal stimulation.," *Clin. Neurophysiol.*, vol. 117, no. 11, pp. 2549–63, Nov. 2006.

[120] N. Kanayama, A. Sato, and H. Ohira, "Crossmodal effect with rubber hand illusion and gamma-band activity.," *Psychophysiology*, vol. 44, no. 3, pp. 392–402, May 2007.

[121] H. GASTAUT, "[Electrocorticographic study of the reactivity of rolandic rhythm].," *Rev. Neurol. (Paris).*, vol. 87, no. 2, pp. 176–82, Jan. 1952.

[122] A. Vretblad, *Fourier Analysis and Its Applications*. Springer Science & Business Media, 2003.

[123] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.

[124] A. Hyvarinen and E. Oja, "Independent Component Analysis : A Tutorial," *Neural Networks*, vol. 1, pp. 1–30, 1999.

[125] P. Tangkraingkij, C. Lursinsap, S. Sanguansintukul, and T. Desudchit, "Selecting Relevant EEG Signal Locations for Personal Identification Problem Using ICA and

Neural Network," in *2009 Eighth IEEE/ACIS International Conference on Computer and Information Science*, 2009, pp. 616–621.

[126] Y. Yao, R. Sun, T. Poggio, J. Liu, N. Zhong, and J. Huang, Eds., *Brain Informatics*, vol. 6334. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[127] C. Taswell, "The what, how, and why of wavelet shrinkage denoising," *Comput. Sci. Eng.*, vol. 2, no. 3, pp. 12–19, 2000.

[128] M. H. Kutner, *Applied linear statistical models*, 4th ed. Chicago: Irwin, 1996.

[129] B. Vidakovic, "Nonlinear Wavelet Shrinkage with Bayes Rules and Bayes Factors," *J. Am. Stat. Assoc.*, vol. 93, no. 441, pp. 173–179, Feb. 1998.

[130] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[131] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *Philos. Mag. Ser. 6*, vol. 2, no. 11, pp. 559–572, Nov. 1901.

[132] M. Aminghafari, N. Cheze, and J.-M. Poggi, "Multivariate denoising using wavelets and principal component analysis," *Comput. Stat. Data Anal.*, vol. 50, no. 9, pp. 2381–2398, May 2006.

[133] B. W. Suter, *Multirate and Wavelet Signal Processing*. Academic Press, 1997.

[134] S. Xie, P. Lio, and A. T. Lawniczak, "A comparative study of noise effect on wavelet based de-noising methods," in *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, 2009, pp. 919–926.

[135] B. R. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE J.*, vol. 44, no. 7, pp. 1596–1610, Jul. 1998.

[136] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[137] and J.-M. P. Misiti, Michel, Yves Misiti, Georges Oppenheim, "Wavelet toolbox," *MathWorks Inc., Natick, MA, 1996*.

[138] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.

[139] G. Buzsaki, *Rhythms of the Brain*. Oxford University Press, 2006.

[140] A. K. Jain, S. Pankanti, S. Prabhakar, and A. Ross, "Biometrics: a grand challenge," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, 2004, vol. 2, pp. 935–942 Vol.2.

[141] N. Kawabata, "A nonstationary analysis of the electroencephalogram.," *IEEE Trans. Biomed. Eng.*, vol. 20, no. 6, pp. 444–52, Nov. 1973.

[142] M. Hangoussi, V. Chrissikopolllos, J. A. Evangciotl, and J. Ij, "Person identification

based on parametric processing of the EEG," in *6th IEEE International Conference on Electronics, Circuits and Systems*, 1999, no. 1, pp. 283–286.

[143] O. Ghitza and M. M. Sondhi, "Hidden Markov models with templates as non-stationary states: an application to speech recognition," *Comput. Speech Lang.*, vol. 7, no. 2, pp. 101–119, Apr. 1993.

[144] M. P. Tarvainen, J. K. Hiltunen, P. O. Ranta-aho, and P. A. Karjalainen, "Estimation of nonstationary EEG with Kalman smoother approach: an application to event-related synchronization (ERS).," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 3, pp. 516–24, Mar. 2004.

[145] D. Fugal, *Conceptual wavelets in digital signal processing : an in-depth, practical approach for the non-mathematician*. San Diego Calif.: Space & Signals Technical Pub., 2009.

[146] A. Jain, *Fundamentals of digital image processing*. Englewood Cliffs NJ: Prentice Hall, 1989.

[147] S. He, R. Dum, and P. Strick, "Topographic organization of corticospinal projections from the frontal lobe: motor areas on the medial surface of the hemisphere," *J. Neurosci.*, vol. 15, no. 5, pp. 3284–3306, May 1995.

[148] D. Y. Kimberg and M. J. Farah, "A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior.," *J. Exp. Psychol.*, vol. 22, no. 4, p. 411.

[149] A. Martin and M. Przybocki, "The NIST 1999 Speaker Recognition Evaluation—An Overview," *Digit. Signal Process.*, vol. 10, no. 1–3, pp. 1–18, Jan. 2000.

[150] NIST Evaluation Tools DETware_v2.1, "DETware_v2.1." [Online]. Available: http://www.itl.nist.gov/iad/mig/tools/DETware_v2.1.targz.htm. [Accessed: 21-Apr-2015].

[151] D. a. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 72–83, 1995.

[152] N. Kawabata, "A nonstationary analysis of the electroencephalogram.," *IEEE Trans. Biomed. Eng.*, vol. 20, no. 6, pp. 444–52, Nov. 1973.

[153] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.

[154] K. Aizawa, Y. Nakamura, and S. Satoh, Eds., *Advances in Multimedia Information Processing - PCM 2004*, vol. 3333. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[155] M. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition," *Speech Commun.*, vol. 54, no. 4, pp. 543–565, May 2012.

[156] S. S. Stevens, “A Scale for the Measurement of the Psychological Magnitude Pitch,” *J. Acoust. Soc. Am.*, vol. 8, no. 3, p. 185, Jun. 1937.

[157] Z. Tufekci and S. Gurbuz, “Noise Robust Speaker Verification Using Mel-Frequency Discrete Wavelet,” pp. 657–660, 2005.

[158] M. I. Abdalla and H. S. Ali, “Wavelet-Based Mel-Frequency Cepstral Coefficients for Speaker Identification using Hidden Markov Models,” vol. 1, no. 2, pp. 16–21, 2010.

[159] S. V.Chapaneri, “Spoken Digits Recognition using Weighted MFCC and Improved Features for Dynamic Time Warping,” *Int. J. Comput. Appl.*, vol. 40, no. 3, pp. 6–12, 2012.

[160] N. E. Huang, Z. Wu, S. R. Long, K. C. Arnold, X. Chen, and K. Blank, “On instantaneous frequency,” *Adv. Adapt. Data Anal.*, vol. 1, no. 2, pp. 177–229, 2009.

[161] E. Huang, Norden E., and Nii O. Attoh-Okine, “The Hilbert-Huang Transform in Engineering.pdf.” 2010.

[162] N. E. Huang, Z. Shen, and S. R. Long, “A new view of nonlinear water waves: The Hilbert Spectrum 1,” *Annu. Rev. Fluid Mech.*, vol. 31, no. 1, pp. 417–457, Jan. 1999.

[163] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009.

[164] J. O. Smith, *Mathematics of the Discrete Fourier Transform (DFT): With music and audio applications*. Julius Smith, 2007.

[165] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Elsevier, 2009.

[166] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” *Am J Physiol Hear. Circ Physiol*, vol. 278, no. 6, pp. H2039–2049, Jun. 2000.

[167] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, “A fully automated correction method of EOG artifacts in EEG recordings.,” *Clin. Neurophysiol.*, vol. 118, no. 1, pp. 98–104, Jan. 2007.

[168] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, Y.-H. Pan, and Y.-H. Wang, “Automatic Stage Scoring of Single-Channel Sleep EEG by Using Multiscale Entropy and Autoregressive Models,” *IEEE Trans. Instrum. Meas.*, vol. 61, no. 6, pp. 1649–1657, Jun. 2012.

[169] “Sample Entropy estimation using sampen.” [Online]. Available: http://physionet.incor.usp.br/physiotools/sampen/. [Accessed: 21-Apr-2015].

[170] S. M. Pincus, “Approximate entropy as a measure of system complexity.,” *Proc. Natl. Acad. Sci.*, vol. 88, no. 6, pp. 2297–2301, Mar. 1991.

[171] D. M. J. Duin, R. P. W., Juszczak, P., de Ridder, D., Paclık, P., Pezkalska, E., & Tax,

"PRTools," 2004. [Online]. Available: http://37steps.com/prhtml/prtools.html. [Accessed: 21-Apr-2015].

[172] J. A. Olvera-López, J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, and J. Kittler, "A review of instance selection methods," *Artif. Intell. Rev.*, vol. 34, no. 2, pp. 133–143, May 2010.

[173] D. Wilson and T. Martinez, "Reduction techniques for instance-based learning algorithms," *Mach. Learn.*, 2000.

[174] N. Network, "Reduction Techniques for Instance-Based Learning Algorithms," pp. 257–286, 2000.

[175] K. Gowda and G. Krishna, "The condensed nearest neighbor rule using the concept of mutual nearest neighborhood (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 25, no. 4, pp. 488–490, 1979.

[176] S. García, J. Derrac, J. R. Cano, and F. Herrera, "Prototype selection for nearest neighbor classification: Taxonomy and empirical study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 417–435, 2012.

[177] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Springer Science & Business Media, 2009.

[178] C. S. Leslie, E. Eskin, A. Cohen, J. Weston, and W. S. Noble, "Mismatch string kernels for discriminative protein classification.," *Bioinformatics*, vol. 20, no. 4, pp. 467–76, Mar. 2004.

[179] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification.," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1055–64, Jan. 1999.

[180] C. Bahlmann, B. Haasdonk, and H. Burkhardt, "Online handwriting recognition with support vector machines - a kernel approach," in *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, 2002, pp. 49–54.

[181] K. Tumer and J. Ghosh, "Estimating the Bayes error rate through classifier combining," *Proc. 13th Int. Conf. Pattern Recognit.*, vol. 2, pp. 695–699, 1996.

[182] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 2000.

[183] and U. S. Beyer, Kevin, Jonathan Goldstein, Raghu Ramakrishnan, "When is 'nearest neighbor' meaningful?.," in *Database Theory—ICDT'99*, 1999, pp. 217–235.

[184] C. W. Anderson, E. a Stolz, and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks.," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 3, pp. 277–86, Mar. 1998.

[185] R. Palaniappan, "Brain Computer Interface Design Using Band Powers Extracted During Mental Tasks," in *Conference Proceedings. 2nd International IEEE EMBS*

Conference on Neural Engineering, 2005., 2005, pp. 321–324.

[186] A. Akrami, S. Solhjoo, A. M. Nasrabadi, M. Reza, and H. Golpayegani, "EEG-Based Mental Task Classification: Linear and Nonlinear Classification of Movement Imagery," in *27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005.*, 2005, pp. 4626–4629.

[187] L. Zhiwei, "Classification of Mental Task EEG Signals Using Wavelet Packet Entropy and SVM 2 Theory Background," in *Electronic Measurement and Instruments, 2007. ICEMI'07. 8th International Conference on. IEEE, 2007.*, 2007, pp. 2–5.

[188] E. Abdalsalam M., M. Z. Yusoff, N. Kamel, A. Malik, and M. Meselhy, "Mental task motor imagery classifications for noninvasive brain computer interface," *2014 5th Int. Conf. Intell. Adv. Syst.*, pp. 1–5, Jun. 2014.

[189] P. F. Diez, V. Mut, E. Laciar, A. Torres, and E. Avila, "Application of the empirical mode decomposition to the extraction of features from EEG signals for mental task classification," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2009, vol. 2009, pp. 2579–82.

[190] J. Sleight, P. Pillai, and S. Mohan, "Classification of Executed and Imagined Motor Movement EEG Signals," 2009, pp. 1–10.

[191] M. Toli and F. Jovi, "Classification of Wavelet Transformed EEG Signals with Neural Network for Imagined Mental and Motor Tasks," in *Kineziologija*, 2013, vol. 45, pp. 130–138.

[192] A. Loboda, A. Margineanu, and G. Rotariu, "Discrimination of EEG-Based Motor Imagery Tasks by Means of a Simple Phase Information Method," *Int. J. Adv. Res. Artif. Intell. (IJARAI), 3(10)*, vol. 3, no. 10, pp. 11–15, 2014.

[193] D. Zhang and A. K. Jain, Eds., *Biometric Authentication*, vol. 3072. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[194] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET Curve in Assessment of Detection Task Performance," *Natl. INST Stand. Technol. Gaithersbg. MD*, 1997.