Increasing Reliability and Security in Handwritten Signature Biometrics

A Thesis Submitted to the
University of Kent
For the Degree of Doctor of Philosophy (Ph.D.)
in
Electronic Engineering

By
Tasmina Islam
June 2017

To my parents and my children

Abstract

Biometric systems are increasingly being used as very convenient and efficient person identification systems, even within large populations. To be able to provide a successful, secure and reliable authentication process, different mechanisms (e.g. developing improved features or classification algorithms or template protection schemes – revocable biometrics, etc.) have been developed to protect systems from attacks and are deployed across many different modalities. The idea of revocability, where a fixed, unchanging biometric template is transformed into a revocable template, has been studied for both physiological and behavioural biometrics. But the concept of "natural revocability" introduced in this study, which most behavioural biometric modalities (and the handwritten signature in particular) offer, provides the possibility of adopting an extremely simple and intuitive strategy for the revocation process without the need for complex mathematical processing, because this is entirely under the user's control. This approach, however, has not been studied in relation to biometrics and data revocability hitherto and the lack of databases to support this type of investigation therefore imposes the need for generation of new data.

The study reported in this thesis investigates the handwritten signature as the target biometric modality of interest in relation to natural revocability by means of an experimental study starting with the collection of "live" signature samples, and wide-ranging subsequent analysis. The suitability and effectiveness of natural revocability in handwritten signature biometrics as a practical option is investigated by observing how "stability" of the form of the signature changes over a period of time. The characteristics of potential revocability are also investigated by analysing performance and invoking the "biometric menagerie" notation for individual behaviour, while a more practically-oriented test of the viability of the natural revocability concept is performed by evaluating recognition performance. A feature-based analysis of the concept is presented by investigating some features

commonly used in signature processing in both longstanding original and naturally revoked new signatures of a group of writers, and exploring the relationship between features, signature style and their effect in relation to original signatures and new signatures. This study also explores the development of a type of feature relating specifically to the concept of hesitancy (or its converse, fluency) for signature processing, which appears to be of particular relevance to the study reported here and investigates its impact on signature development in the context of natural revocability and signature verification more generally, using an objective measure of the power of the feature.

The results from the experiments and the analysis provided suggest that if a sufficient time period is allowed then there is a high likelihood of convergence in terms of stability between a highly practised and long-standing signature and an alternative new representation, which also can be reliably recognised without degrading recognition accuracy, supporting the suitability and viability of the natural revocability concept. Exploring the influences of objectively defined hesitation features in creating the new signature also reveals that signers are more hesitant initially in signing the new signatures than the original - as might be expected - but gradually the hesitancies reduce with time, showing signers' increasing confidence in signing new signatures as time progresses; similarly, investigating recognising genuine and forged signatures shows that the hesitation is higher in forgery signatures than in genuine signatures, supporting the qualitative definition of hesitancy applied in a typical forensic scenario, an encouraging and effective performance improvement in discriminating genuine and forgery signatures. The study reported brings together two related ideas: the possibilities of adopting a natural revocability strategy in relation to security and reliability in handwritten signature analysis, and the development of a feature which may be particularly effective in the area of handwriting analysis, together with the aim to throw new light on issues relating to security in practical biometric systems.

Acknowledgements

All praises are for the Almighty.

At first, I would like to express my warm thanks to my supervisor Professor Michael Fairhurst for his continuous guidance, encouragement and patience throughout the Ph.D. research and writing of this thesis. It has been a privilege and pleasure to work with him as one of his Ph.D. students and I will never forget his unfailing support and inspiration even during difficult times and the time spent discussing various ideas.

I gratefully acknowledge the financial support from the School of Engineering and Digital Arts, University of Kent, UK for this research work.

I would also like to thank all the members of staff in the School of Engineering and Digital Arts. I am particularly grateful to Dr. Farzin Deravi for his cooperation during difficult times and Jo and Yan for their prompt help in sorting out IT problems.

Many thanks to my friends and colleagues, Dr. Márjory Cristiany Da Costa Abreu, Dr. Meryem Erbilek, Dr. Richard Boner, Dr. Asad, Dr. Cheng, Iulia Motoc, Hamed Alsufyani, Nawal Alsufyani at the Intelligent Interaction Research group of School of Engineering and Digitals Arts of the University of Kent for making my life more enjoyable.

Finally, I would like to express my heartfelt gratitude to my family: especially, to my mother Suraya Begum, my father Md. Tajul Islam, my sister Sadia Islam, for their never-ending love, care and support; to my husband, Dr. Hannan Bin Azhar, for his continuous support and inspiration especially during the write up of this thesis; and to my boys, Suzerain and Ruzain for making me smile even in difficult times and understanding me.

Table of Contents

List of Figures

List of Tables

Chapter 1:

Reliability and Security in Handwritten Signature Biometrics

Security and reliability are very important issues for biometric systems if they are to be able to provide a successful authentication process while protecting from attacks. Different mechanisms (e.g. developing improved features or classification algorithms or template protection schemes, etc.) have been developed to counteract such attacks (e.g. providing revocable biometrics) and are addressed for different modalities. But relatively little work has been reported in the area of behavioural biometrics, such as the handwritten signature. This thesis will describe a study to explore this issue in a new and especially natural context, exploiting the particular characteristics (e.g. the possibility of adopting an extremely simple and intuitive strategy for the revocation process) of such behavioural modalities, and the handwritten signature in particular. This study will investigate the handwritten signature modality to provide a practical environment for experimentation and analysis.

This first chapter will present the fundamental background and basis for the investigations and analysis reported in this thesis. Section 1.1 will describe the general motivation of the thesis. Subsequently, Sections 1.2 and 1.3 will report an overall review of biometric systems, relevant security concerns, and the protection mechanisms commonly discussed in the literature, providing a context for the new work to be reported later in the thesis. Section 1.4 will provide a review of the current state of the art of the handwritten signature modality. Following this Section 1.5 will outline the principal contributions of this thesis to the field, and finally the organisation of the study to be presented in this thesis will be explained in Section 1.6.

1.1 General motivation

Biometric authentication is a process of determining or confirming an individual's identity through a measurable biological characteristic (which can either be anatomical or physiological, for example, fingerprints, iris patterning, hand geometry, etc.) or a behavioural characteristic (handwriting/signature, voice characteristics, keystroke patterns, gait, etc.) [1]–[4]. Some examples of commonly used and less commonly used biometric modalities are shown in Table 1.1. A physiological biometric such as a fingerprint pattern can be imaged by locating the end of the finger, while a behavioural biometric, for example, handwriting or a handwritten signature requires the user to execute a writing activity before the biometric data (handwriting or signature) is available.

Table 1.1 Some common and less common biometric modalities

Physiological	Behavioural
Face	Signature/Handwriting
Fingerprint	Voice
Hand geometry	Gait
Iris	Keystroke
DNA	Lip motion
Ear shape	
Odour	
Retina	
Skin reflectance	
Thermogram	
EEG	
ECG	
Palm print	

A biometric system has been shown to be far more efficient and effective than a more traditional token- or password-based authentication system which relies on the person's knowledge or belongings rather than who the person is. Therefore, biometric systems are increasingly being used in border control environments or in airports [5], justice and law enforcement applications (for example, criminal identification, forensic investigation in court cases, identification of a missing person), financial/transactional support (for example, mobile or internet banking), physical and logical access controls, healthcare and so on [1], [2], [6], [7]. As a result of their wide deployment, biometric systems, like any other systems are vulnerable to deliberate attack by an adversary [8]–[12]. Thus, the security and reliability of biometric systems play a very important role in the biometric authentication process and different protection mechanisms have been reported to counteract these issues [13]–[16].

Among all biometric modalities, handwritten signatures have been the most socially and legally accepted means for identification [17], [18]. Though the signature is notorious (as a behavioural biometric) for its sometimes-high intra-sample variability, practising the writing of the signature over time generally creates a natural recognisable pattern which then can be used in identity verification. Unlike physiological biometrics, such as the fingerprint (as noted above), signatures cannot be produced unintentionally or while the user is unconscious, and therefore the signature is likely to be retained in many identity verification tasks. Being a behavioural biometric the handwritten signature is claimed to possess some important natural and inherent characteristics, some of which have often been overlooked. The motivation of the study reported in this thesis is to investigate and explore some of these characteristics in increasing both security and privacy, but also avoiding the need for developing alternative and more complex protection mechanisms while offering opportunities to increase significantly the attractiveness and reliability of handwritten signature as a biometric modality. This can have implications not only in the field of biometrics itself but also in a wide variety of application areas (for example, health care, forensic scenarios, etc.).

However, first of all, it is important to understand the current situation with regard to security and reliability issues in biometric systems. Hence, the remaining sections of this chapter will first give a brief overview of biometric systems in general and associated security concerns and protection mechanisms with respect to several different key biometric modalities; then an overview of the handwritten signature biometric modality in relation to the signing process, signature authentication systems and its use in different disciplines will be presented, identifying and exploring issues relevant to the study which still need further investigation and clarification. Finally, the chapter will identify the main contributions made through this study, and will describe the organization of the further chapters of this thesis.

1.2 Overview of biometric systems

Generally, a basic biometric authentication system, whether based on either physiological or behavioural modalities, operates in two main stages: enrolment and authentication. In the enrolment stage, biometric data of the chosen biometric modality are captured from an individual (a person) using a sensor or capture device, then feature extraction is performed to extract appropriate features from these raw data and using the extracted features a model is constructed representing each individual (with some form of ID linked to these features along with other data such as the person's name), which can then be stored as biometric templates (one for each enrolled individual) in a database for future authentication purposes. In the authentication stage an individual is either identified or verified by the biometric system. The *identification* process involves establishing a person's identity through searching an entire database, whereas the *verification* process involves the authentication of a person's claimed identity by comparing or matching the person's acquired biometric information to that of the stored template of the person's claimed identity. The processes of enrolment and authentication are carried out using several processing steps and the design of these steps can differ significantly depending on the biometric modality adopted. However, most biometric systems require a processing chain encompassing the following five main steps:

- **Acquisition**: is the step where relevant data for the biometric modality of interest are captured using a biometric sensor or capture device from an individual. For example, if the modality of interest is the fingerprint, then a fingerprint image must be captured at this stage. This is one of the critical modules, as the performance of the entire system is often affected by the amount of care taken in the data acquisition [4], [19], [20].

- **Pre-processing**: is the step where the acquired data are processed according to the type of the biometric modality and prepared for the further steps. Depending on the biometric modality, this step can be carried out either before or after the feature extraction. For example, this step can be simply a normalisation of the acquired values from the signature biometric modality, but may include various image enhancement processes, and so on.

- **Feature extraction**: is the step where appropriate discriminative identity-defining information (features) are extracted from the acquired raw or pre-processed biometric data. A *feature selection* step is often used after the features are extracted to select the most important or effective features for use in the given application.

- **Matching**: is the step where the extracted features during the authentication stage are compared against the stored templates generated from the extracted features during the enrolment stage to generate matching scores. Thus, a test sample will generate a score which reflects the similarity between that sample and the stored template of interest. Following this, decision making is performed by using the matching scores based on a threshold to confirm/deny an individual's claimed identity (verification) or to identify the user (identification) as one of the set of enrolled users. For example, the identity of the owner of the test sample might be judged to be the individual whose template generates the maximum score across all the templates.

- **Database**: is where the models of the enrolled individuals (constructed from the extracted features) are stored as templates (a compact representation of the acquired biometric sample is known as a biometric template [21]) for future authentication purposes.

1.3 Biometric security

Biometric systems, like most systems, are also vulnerable to potential security threats leading to adverse consequences such as repudiation claims by corrupt users (users deny accessing the system claiming their data were acquired covertly or circumvented [22]), legitimate users being denied (denial-of-service), unauthorized users intruding the system(by spoofing) and so on. Figure 1.1 shows some potential adversary attacks and attack points on biometric systems.

Many of these attacks are applicable to any authentication systems such as replay (where an adversary replays a stolen electronic copy of the stored biometric data to the system avoiding the biometric sensor), Trojan Horse (where the feature extractor, matcher or the decision module is substituted by a Trojan Horse program to generate either desired feature sets or match scores or decision irrespective of the identity of the user presenting the biometric sample at the sensor), man in the middle attack (where an adversary intercepts the data while it is being transported through the channel and inserts a modified version of the data back into the system). But there are two adversarial attacks which are more specific to biometric systems. Spoofing is one of them [23]–[26], where a fake biometric generated typically by covert acquisition of a genuine individual's biometric data, is presented at the biometric sensor to access the system [2], [21]. For example, lifting latent fingerprint impressions from objects touched by an individual and using these to construct a mould which is later used to gain unauthorised access to the biometric system [22]. Another major threat is attack on the template, where the stored template (stored either locally or centrally) generated during the enrolment stage, is modified or replaced by a new one. Typically, biometric information is permanently associated with its owner and if fraudulently obtained or simply stolen , this cannot be replaced [14], [27]. In the event of compromise, the attacker can generate fake samples to gain unauthorised access to a biometrically-protected system [28]–[32]. If the same biometric feature is used by a user in multiple applications, then the attackers can use the same mechanism to compromise all applications, and furthermore if organisations share biometric data, they (attackers) can potentially track the user.

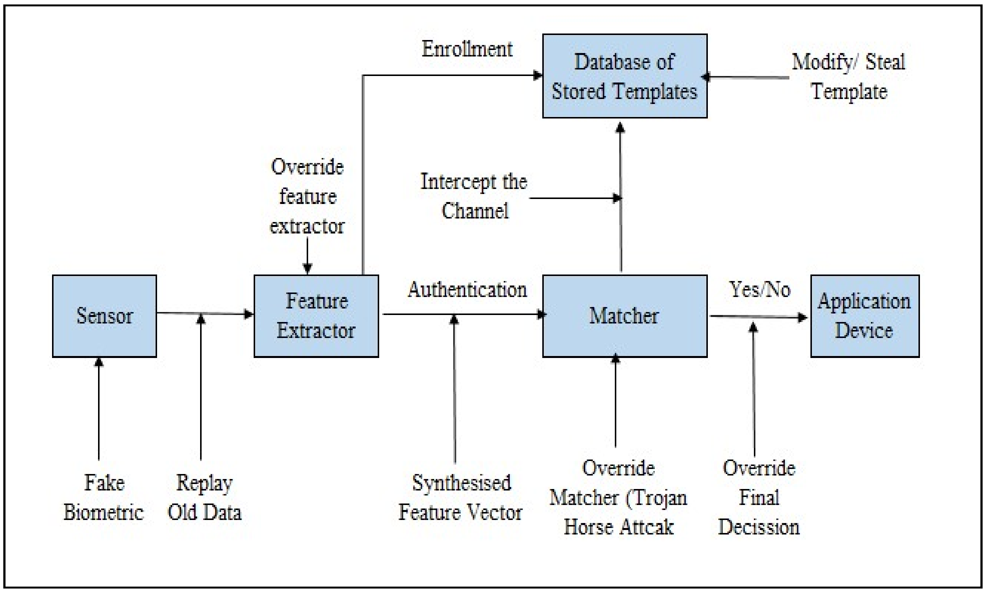


Figure 1.1 Possible attacks and attack points on a generic biometric system, adapted from [9]

When a biometric system is used as part of a larger scale security system, some other vulnerabilities associated with larger system requirements can be observed

[33], [34] such as damaging the biometric sensor or disabling the power supply can cause the whole biometric system to become ineffective [9], [35] or even sometimes non-biometric modules that are part of the overall security system can reduce the effectiveness of the biometric system [36]. Several factors, such as template corruption due to an accidental system failure, template alteration by deliberate adversarial attack or template substitution (replace a valid one with a fake template) to hinder the system, are listed by the UK Biometric Working Group (UK-BWG) [37] by which the quality of the template can be affected.

Intentional alteration of biometric traits (for example, use of altered fingerprints) to avoid identification (for example, individuals want to avoid identification because of their previous criminal convictions) is also a major security concern in some applications (for example, international border control) [38]. In some deployed biometrics systems, fraudulent samples (such as static iris images [39], static facial images [40], [41]) and fake fingerprints [42]) presented at the biometric sensors, are processed as genuine biometric samples collected from genuine users and sometimes are verified successfully by matching against the stored templates. Thus, the detection of spoofing has become a critical requirement of biometrics systems. Spoof detection is usually performed using techniques which typically check for signs of human “liveness” (such as fingerprint sweat, blood pressure, detection of a pulse, or specific reflection properties of the eye) giving rise to the commonly accepted term *liveness detection* [43] as a collective term for a range of such techniques. A range of approaches have been reported in [41], [44]–[48] in detecting altered and/or spoof biometrics samples.

In order to overcome the vulnerabilities of biometric systems, several template protection schemes have been proposed in the literature [2], [9], [21], [27], [49]–[57]. According to [8] the concept of *cancellable* biometrics [27], [58], also known as *revocable* biometrics (of particular interest in our present context), has been introduced as one of the protection schemes and has been paid a lot of attention in the research community in recent years [13], [16], [27], [59]–[63]. In this method, the fixed unchanging biometric template is transformed using a unidirectional function into a 'revocable' template. The underlying principle here is that, in the event of compromise of the template data, the biometric data can be re-enrolled using another different transformation, while the original biometric information remains protected and unchanged. Because of the properties of the unidirectional transformation which is adopted, it makes the retrieval of the original biometric data extremely difficult from the transformed protected template, providing security of the template and, as a different transformation is applied in a different application it prevents cross-matching between databases, preserving the privacy of the person who owns of the biometric data. It is also shown that the matching performance does not seriously degrade since the statistical attributes are almost maintained when constructing the biometric template in this way[13]. It should be noted that this cancellable or revocable biometric template protection scheme is closely related to cryptosystem-based scheme, but not equivalent. As described earlier, in the revocable template protection scheme computationally recovering the original template is extremely difficult from the transformed template because of unidirectional transformation even if the secret key is made available, while some information about the template is made available (also known as *helper data)* in the cryptosystem-based scheme to generate cryptographic keys during matching process. A detailed review of biometric cryptosystems can be found in [9], [64]–[66]; some examples of biometric cryptosystems in the literature include fuzzy vault [49], [54], distributed source coding [67], shielding functions [68], fuzzy extractors [69] and fuzzy commitment [70].

Different strategies for generating revocable biometric templates have been proposed in the literature. In a recent review of cancellable biometrics published in [15], the authors have broadly divided the available cancellable biometric template protection schemes into two categories: i) methods that need a particular matcher to work with and ii) methods that do not need a particular matcher (i.e. able to work with available conventional matchers). Examples of methods requiring a particular matcher include, Biotokens [71], BioConvolving [61], salting [72], PalmPhasor code [73], correlation-based MACE filter approach [74]. Whereas methods such as combo [72], block-remapping [75], image warping [75], non-invertible transforms [13], [58], [76], dynamic random projections [77], permutations [78], random projections [63], BioHashing [79], [80], PalmHashing [81], Minimum distance graph [82], curtailed circular convolution [83] can work with available conventional matchers. Most of these methods need good registration of the biometric samples, for example BioConvolving, non-invertible transforms, salting, and some are registration free methods such as correlation-based MACE filter approach, minimum distance graph. Some of these methods (e.g. Biotokens, block-remapping etc.) work with the original biometric samples where some, for example, BioHashing, permutations, work with extracted features (e.g. wavelet transforms [81], Gabor features [72]) from the original biometric samples. These different revocable biometric schemes have been implemented for a number of biometric modalities including the following; fingerprint [13], [71], [77], [79], [82], [83], iris [63], [72], [75], palmprint [73], [81], face [78], [84], and signature [61]. A further detail review of these revocable techniques applied to different biometric modalities will be discussed in Chapter 4. As can be seen from the many studies reported in the literature, this idea of revocability (revoking a new template through transforming the original template) has been studied for both physiological and behavioural modalities. It is evident that most behavioural modalities, and the handwritten signatures in particular, present the possibility of adopting an extremely simple and intuitive strategy for the revocation process. Since such modalities depend entirely on an action carried out by an individual rather than an inherent physiological characteristic, the revocation process can be entirely under that individual's control. We have called this approach "natural revocability" (a detailed explanation and discussion is given in Chapter 4) and clearly, this is an area which will benefit from further investigation.

1.4 Handwritten signature

The handwritten signature has long been established as a common means for providing proof of identity especially in financial transactions, certification of documents, wills and other legal documents, etc. Because of its widespread use, signature checking based authentication is very familiar and highly acceptable to the public, hence putting the handwritten signature in an advantageous position over other biometric modalities in relation to suitability for a range of practical applications.

1.4.1 Signing process and signature development

The handwritten signature (as a specific manifestation of general handwriting) is the product of a learned neurophysiological motor program which is a complex interaction of cognitive and neuromuscular and biomechanical [85], [86] processes. This fine motor act involves (Figure 1.2) the brain which initiates the motor control via stimuli; the peripheral nervous system which modulates the movement by relaying impulses from the spinal cord to muscles; the muscles themselves which then activate the appropriate limbs (arm/ wrist/ fingers) by contraction or relaxation in a coordinated fashion; and finally, the limbs which execute the signature using a writing device (e.g. pen, paper etc.). Quoting from [86] - "according to this model, some central nervous system mechanisms within the brain fire, with a predetermined intensity and duration, the nerve network which activates the proper muscles in a predetermined order. The motion of the pen on the paper, resulting from muscle contraction/relaxation, leaves a partial trace of the pen-tip trajectory".

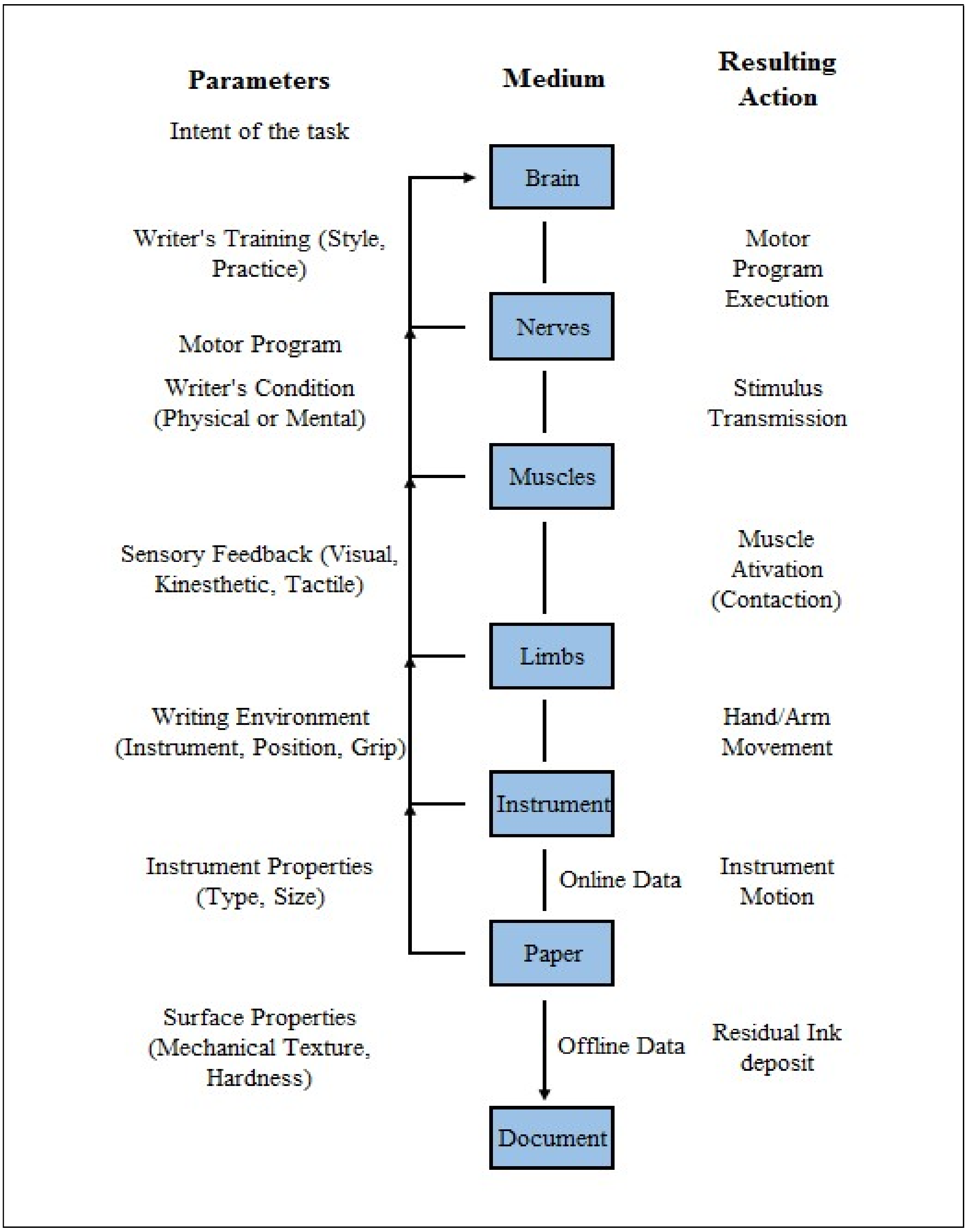


Figure 1.2 Schematic overview of the writing/signing process adapted from [87]

The signing process is derived from the normal, more general handwriting process and starts with lines and scribbles at a very early age. In the pre-school (between age 3 and age 4) stage, children begin to write letters and numbers, usually copying or tracing from printed worksheets. This helps to create the spatial memory or cognitive map by improving their motor control. This also helps the children to learn different shapes and understand the spatial concept of objects in relation to other objects [88]. Once the motor skills begin to mature (by practising over time [89]) the person starts to write fluently and effortlessly and at this stage s/he develops his/her own signature style [90].

1.4.2 Signature variation

Due to the nature of the signing process, appearances of two samples of the same signature written by the same writer are never absolutely identical. Thus, there is always a variation between samples of a person's signature, which is also referred to as *intra-personal variability*. This variation can be influenced by the signing condition, such as the position of signing and pen grip as well as the writing instruments and writing surface used [91], [92] , the psychological or environmental (e.g. stress) [93], [94] or physical condition [95]–[100] of the signer at the time of signing, ageing of the signer, and so on. This intrinsic variability of signatures is an important and sometimes subtle factor that should be taken into consideration in the design of automatic signature verification systems, as this highly affects their performance.

As mentioned earlier, the signing act involves a complex interaction between the brain, the central and peripheral nervous systems as well as the muscles and limbs (Figure 1.2), and any influence on these parts of the motor system will affect the writing and signing behaviour. The genetic makeup of these parts also contributes to individual writing or signing style (which leads to a natural *inter-person variability*). For example, important factors include the relative shapes, sizes, and locations of the hand bones (for example, carpals – wrist bones, metacarpals and finger bones) influencing the writer's pen grip; whether the writer is right or left handed; the strength of the muscles, peripheral nervous system disorders (such as peripheral neuropathy resulting loss of coordination and muscle weakness) etc. The authors of the study reported in [101], [102] refer to these as *genetic* (biological) factor, and identify them as one of the fundamental factors that contribute to the individuality of the writing style. Another factor, described as *memetic* (cultural), is the cultural influence on how a person grips the pen, writes the characters or letters (allographs). Usually people learn these during education or by observing other people's writing or from the printed worksheets employed. As the learning strategies are quite diverse, writing style differs not only from one country to another but also between schools [103], [104]. Analogous to handwriting style, signature style is strongly influenced by cultural habits and languages. For instance, in the USA a signature is usually executed in neat and legible handwriting, where European signatures can commonly have more of a resemblance to more abstract shapes, where the main body of the signature is usually surrounded by a number of curves, lines, and loops. In order to consider a signature as legally relevant in Germany, at least three letter shapes have to be legible, where a Persian signature is more of an ornament that is barely legible. Furthermore, Japanese and Chinese signatures are made up of independent symbols, while Bangla and Hindi signatures do not have any case concept (upper or lower case) and people write Arabic signatures starting from right to left as opposite to most of the other signatures. Therefore, for an automatic signature verification system to be universally successful, it should incorporate design considerations addressing these differences. Examples of some specific approaches proposed in some studies reported in the literature for single script include Chinese, Japanese, Arabic, Persian, Bangla and Hindi signatures [105]–[109] and for multiscript include English and Chinese, Dutch and Chinese, English and Hindi [110]–[113] etc.

Due to these different kinds of factors that affect the signature characteristics, it can also be important to consider a set of metadata, also known as *soft biometrics* [114], related to some personal information about the signer such as, gender, ethnicity, handedness, script language, age etc. The study reported in [114] defined soft biometrics as follows: "Soft biometric traits are those characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals". Some studies reported in the literature on soft biometrics based on signature biometrics can be found in [56], [115]–[118].

1.4.3 Signature based authentication

Like the basic biometric system described in Section 1.2, a basic signature authentication or verification system is also operated using the following modules or stages: acquisition, pre-processing, feature extraction, matching and a database to store signature templates. In general, signature verification can be divided into two categories based on the acquisition method adopted, i) static, also referred to as *offline* capture and ii) dynamic, also often referred to as *online* capture [17], although online capture can also give rise to dynamic features. In the offline mode, the signature is acquired when the writing is completed, generally through a camera or an optical scanner and represented as a digital image, usually in grayscale format, as a set of points $S(x, y)$, $0 <= x <= W$, $0 <= y <= H$, where W and H are the width and height of the image. In the online or dynamic mode, signature is generally acquired using a graphics/digitising tablet with an electronic pen while the signature is being executed. The data in online mode is collected as a time sequence $S(n)$, where $n = 1, ..., N$ is the discrete time index and $S(n)$ is the signal sampled at time $n\Delta t$, where Δt being the sampling interval. Depending on the capture device, the signal may contain only the position data (horizontal x and vertical y position) of the pen at each sampling point, or may include additional data such as the pen inclination, pressure, etc. [119]. However, in this mode, it can be seen that information can be captured not only about the overall appearance of the signature (image), but also about the execution of the signature – that is, time-based information.

Digitising tablets with electronic pens are the most traditional online signature acquisition devices [120], [121]. Some of these devices use touch screen display with digital-ink technology that provide immediate visual feedback to the writer while signing (e.g. tablet PC) [122], [123]; some devices use a standard paper overlaid on the tablet surface with an inking pen, which not only allows the writer to write conventionally using pen and paper but also produces an exact digital copy of the actual handwritten signature, allowing both offline and online acquisition of the biometric data at the same time [3], [124]. In addition to standard electronic tablet and scanner for online and offline signature acquisition, other devices have been used, for example, personal digital assistants (PDAs) and smartphones[125]–[128]. Studies reported in recent years have also used camera-based acquisition systems [129], [130] by recording sequential pen tip tracking using a webcam for online signature verification. Another study reported in the literature proposed a video-based acquisition system for in-air signature (writing virtually in front of the camera) [131]. Although, some devices raise specific issues, such as writing with a fingertip or stylus instead of a pen or using a touch screen based surface (or even in the air) instead of writing on paper or collecting a limited number of data or poor sampling frequency of the collected data, etc., the increase in the possible number of acquisition devices will make signature acquisition and eventually automatic verification of signature more feasible and usable in various applications. Figure 1.3 shows some acquisition devices reported in the literature.

A pre-processing step is generally performed, with the aim of reducing noise as much as possible in order to obtain the most accurate signature representation for later authentication purposes. Though there may be an argument that potentially useful information may be lost in this process [132], it can affect the other successive steps of the authentication process [119] and in some cases it is crucial to perform this step. Common pre-processing approaches include filtering, noise reduction, normalising, segmenting and smoothing. Details of the typical algorithms for filtering, noise reduction, normalising and smoothing can be found in [133]–[138]. Segmentation of the signature can be a very complex task because of the intra-person signature variability and is a crucial pre-processing step especially for offline signatures. The most common offline signature segmentation techniques are based on an analysis of connected components [134], [139], analysis of tree structures obtained from projection profiles (horizontal and vertical) [140], and statistical analysis of directional data [141]. Some online signature segmentation approaches are based on specific signature information (such as pen pressure) and simply derived from the collected signature signal by considering the signature as a sequence of writing blocks (when the pen is down) and interruptions (when the pen is up) [142]. Other approaches are based on the analysis of velocity signals [143], dynamic time warping (DTW) [144], and so on.

Figure 1.3. Signature acquisition devices [3], [130], [131], [145], [146]

Features used for handwritten signature processing can be broadly divided into two types: i) features where signature is represented as a string of time-based signals or functions, also referred to as *function features* [119], and ii) features where signature is represented as a vector of some specific measures of signature characteristics, also known as *parameter features*. The parameter features can further be divided into two types: a) global features (where extracted features represent the whole signature) and b) local features (where extracted features represent specific parts of the signature). Generally, function features are considered for online signature verification and provide greater verification performance compared to parameter features, but the matching process usually requires a long time [119]. Examples of online function features include pen tip position (also applicable in offline), its moving direction, velocity, acceleration, inclination or angle, pressure [147]–[151]. Global parameter features used in the studies reported in the literature typically include total signature execution time, number of pen ups or pen downs, different measures of position, velocity or acceleration (e.g. average or maximum or minimum, etc.), time duration of positive or negative position, displacement, velocity or acceleration etc. for online where size and shape related features are also used in offline [152]–[155]; and various statistics-based features computed by applying mathematical transforms, such as Fourier and Wavelet transforms [156]–[158] for both online and offline, and the discrete cosine transform for online [159] signature verification. Studies reported in [18], [134], [140], [160], [161] use grid-based, orientation-based, geometric-based, structure-based and projection-based, and local features for offline signature verification.

The efficiency and cost of a signature verification system and other processing requirements depend on which and how many features are used in the feature set for the verification, and therefore feature selection is also an important aspect of a signature verification system [162]. Various feature selection approaches have been proposed in studies in the literature, primarily based on Sequential Forward /

Backward Search (SFS/SBS) [163], Principal Component Analysis (PCA) [164], Inter-Intra Class Distance Ratios [165], Neural Networks [166], Self-Organizing Feature Maps (SOFM) [167], Genetic Algorithms [168] and Gain-Ratio Attribute Evaluation [169].

Like the basic biometric system described earlier in Section 1.2, in the authentication process a test signature of a user is either verified as the person s/he claims to be or identified as one of the enrolled users by comparing its extracted features against the stored templates generated from the extracted features during the enrolment stage, which generates a matching score. For verification, matching is performed between the features of the test signature sample and the stored templates of the claimed identity, while for identification, it is performed between the test sample and all the stored templates. Matching score generated from the matching step is used to arrive at a decision based on a threshold to confirm/deny an individual's claimed identity (verification) or to identify the user (identification).

Various approaches are adopted in the matching process for both offline and online signatures and according to [119] are initially divided into three main approaches: template matching, statistical and structural. The most common template matching approaches used for online signature verification are based on Dynamic Time Warping (DTW) when function features are used. Because of the possibility of occurrences of a writer's hesitations or pauses, deletions, additions or gaps in the signature signal sequences, this can complicate the matching process. DTW finds and obtains an optimal match by allowing expansion and compression of the time axis of signature signal sequences, which makes it suitable for online verification [170], [171]. Other template matching approaches include fuzzy logic, relaxation matching, split-and-merge mechanisms, displacement functions [172]–[174]. A statistical approach based on using Support Vector Machines (SVM) has been used successfully in both online and offline signature verification [109], [175]. Hidden Markov Models (HMM) have also been used often because of the conformability of the technique to intra-person variability [148]. Distance-measure based approaches such as Euclidean distance, Mahalanobis distance are preferred when parameter features are used [149], [176]. Neural Networks (NN) in signature verification have been used for a long time, for example the Multi-Layer Perceptron (MLP), Bayesian networks etc. [177], [178]. Structural approaches, for example, tree, graph and string matching techniques in combination with other techniques have been used in studies reported in the literature [170], [171], [179]. Various Multi-Classifier Systems (MCS), combining classifiers, have also been proposed for online and offline verification [180]–[182]. These systems can outperform individual local and global feature based verifiers, because the majority of these systems use a combination of both global and local features.

There are two types of error associated with automatic signature verification, i) false rejection of genuine signatures, also referred to as *Type I error* and ii) false acceptance of incorrect or forged signatures, referred to as *Type II error.* Type I and Type II errors are calculated as false rejection rate (FRR) and false acceptance rate (FAR) respectively; where the FRR is calculated as the percentage of falsely rejected genuine signatures out of the total number of genuine signatures shown and similarly the FAR is calculated as the percentage of forgeries falsely being accepted out of the total number of forgeries shown. Generally, these two error rates are used for evaluating the performance of a verification system. A decrease in one error rate leads to an increase in the other, therefore, the selection of a decision threshold in matching depends on the requirement of the application in to minimise the appropriate error rate. Often, to evaluate the overall error rate of a verification system, an equal error rate (EER), when FAR is equal to FRR, is used.

1.4.4 Applications of handwritten signature biometrics

Due to the nature of handwriting and the writing/signing process this biometric modality has been studied in many disciplines and is still a very active ongoing research area. As noted earlier, the signing process is a complex and fine motor act which requires interaction between cognitive and neuromuscular and biomechanical processes that generates a fast, fluent handwriting movement to produce the learned product, handwritten signature. Any changes in the usual signing behaviour or natural variability may give an indication of changes in neurophysiological or cognitive properties and is very useful in medical diagnosis [183], [184]. This idea has been applied to detect neurodegenerative diseases by analysing some simple handwriting tests carried out by the patients. Examples include Alzheimer's, Parkinson's or Huntington's diseases, spinocerebellar ataxia (which badly affects the hand-eye coordination) and Friedreich's ataxia (which damages the nervous system) [97], [98], [185]–[188]. For instance, in a study reported in [189], the effect of rapid muscle contraction or relaxation of muscles during handwriting movement is found to provide significant information about neurodegeneration. Another study reported in [100], shows that degradation of handwriting can help in diagnosing the stage of Alzheimer's disease in patients. An analysis of the variability of grip kinematics during handwriting reported in [190] shows that it can be useful in diagnosing diseases at a preliminary stage. The rapid movements of the hand during the signing process have been studied extensively [126], [191], [192] and the velocity profile model of these movements are currently being used in identifying risk factors for prevention of brain strokes [183], [193]. Handwritten signatures have been analysed to estimate or predict emotions, intelligence, social attitude, personality and social skills as they are said to subconsciously reflect the personality of the writer [194]–[196]. Above all, handwritten signature analysis plays a key role in validating documents such as contracts, testaments, corporate tax returns, government legislation, etc. Usually this validation is carried out by professional forensic document examiners (FDE) mostly focusing on visually detectable features (e.g. unnatural pen lifts, hesitation, tremor etc. [197]) which in handwritten signatures form the basis of evidence supporting whether a questioned signature is genuine or forged. Automatic signature verification can provide substantial support in the field of forensic analysis, and there is an increasing trend towards bringing together researchers in biometrics and in forensics. Researchers have reported various automatic signature verification systems over the last few decades [119], [198], but most of these systems are developed either for general purpose or considering specific industry requirements, or cannot be directly applied to a forensic scenario (sometimes due to the differences in discriminating characteristic terminologies used in forensic and automated signature analysis). Although, according to [7], [199]–[201], the work of professional FDEs have a great influence on the research carried out in automatic signature verification field as, it has not been possible to computationally extract every feature utilised in forensic document examination [202] and sometimes this has been overlooked. Therefore, this remains an issue in the signature analysis field in terms of its applicability and reliability, and is a topic to which we shall return in the present study.

So, there is clearly a demand for objective and reliable testing procedures in distinguishing authentic signatures from forged signatures. In the study to be reported here, two of the areas referred to in this initial review will be of principal concern. The first concerns the investigation of the idea of what we have termed "natural revocability in the handwritten signature, as an example of a widely used biometric modality. The second is the development of features which may be particularly effective in the area of handwriting analysis. These two ideas are closely related, and the present study will bring these together to throw new light on issues relating to security in biometric systems.

1.5 Contributions

The key point emerging from the review and discussion presented so far is that security and reliability of biometric systems is an ongoing issue, especially in behavioural biometrics, which we suggest can be seen to offer the effective adoption of natural revocability, which has not been studied in relation to biometrics and data revocability hitherto. Thus, the main contribution of this thesis is the investigation of natural revocability which may open up possibilities for revocability strategies proving to be both simple and effective, offering opportunities to increase significantly the attractiveness and reliability of the handwritten signature as a general-purpose biometric modality. More specifically, this thesis contributes the following novelties to the field.

- A novel signature database is developed collecting samples of a newly developed signature for each subject as well as subject's original signature over a significant number of capture sessions (four, six, and ten) to study the nature of natural revocability in the handwritten signature. No other such database exists.

- A new concept of "natural revocability" for handwritten signature is defined and investigated (in Chapter 4).

- The practical viability of the concept of natural revocability when replacing a longstanding signature (original signature) in terms of gaining stability with time is investigated and reported, and we present evidence to show that attaining stability in a relatively short time scale (for most of the signers) is achieved. Also, the variation in stability is investigated from the longstanding original to naturally revoked new signature using the nomenclature of the "biometric zoo/menagerie". This is one of the first studies of online signature revocability coupled to the use of biometric menagerie terminology (described in detail in Chapter 4).

- Variation in signature features and signing style when changing from longstanding original signature to new signature following revocation is investigated.

- A new algorithmic feature to objectify a hitherto qualitative concept relevant to handwritten signature processing is developed and its impact on signature processing investigated.

- Human detection of genuine and forgery in the context of natural revocability is investigated in a preliminary study.

1.6 Chapter conclusion and thesis organisation

This chapter has presented the general motivation of the study reported in this thesis, with an overview of the previously reported research related to the study reported in the thesis, in order to provide a clearer view of the current state of the art. However, each of the subsequent chapters will review the relevant material in a more comprehensive, detailed and focused way in order to illuminate relevant issues relating to each specific analysis to be performed. In the light of the initial reported review, some specific research problems have been identified, where a better or more extensive exploration and investigation is necessary if an enhancement in the reliability and security of practical biometric systems is to be achieved. Subsequently, the specific contributions made to the field from the study to be presented in this thesis are reported.

Finally, in order to make the following chapters more cohesive and easier to follow, it is useful to set out and draw attention to the organisation of this thesis as follows:

Chapter 2: Experimental infrastructure

This chapter will provide an overview of the basic experimental infrastructure and the important practical details used in all the experiments reported in this thesis. The details of data acquisition, feature extraction and some feature processing (respectively), and classification software used in the experiments and analysis for this study will be presented. Also, a discussion of the available online signature databases, which the literature shows to have been commonly used previously for signature studies, as well as the databases and datasets utilised in carrying out the experiments (to be described in greater detail in subsequent chapters) for this study will be presented. Beginning with this overall consideration of the basic tools and experimental infrastructure will make the subsequent chapters, which report and analyse a range of experimental studies, more cohesive and easier to follow.

Chapter 3: Revocability database compilation

This chapter will discuss the handwritten signature data collection protocol and ethical approval procedure, together with an overview of the acquisition system. Also, an illustration of the collected data and acquired subject information will be presented as well as a discussion on the challenges faced during data collection. The long-term value of this database will be explained.

Chapter 4: Natural revocability in handwritten signature biometrics

This chapter will present the experimental analysis of the handwritten signature samples collected to support the initial study of our concept of natural revocability in handwritten signature biometrics. A review of the revocability studies reported in the literature in different biometric modalities, together with a discussion on the idea of natural revocability and its viability in handwritten signature recognition will also be presented.

Some of the work to be reported in this chapter is published in [152]

Chapter 5: Feature Based Analysis of Natural Revocability in Handwritten Signatures

This chapter will present some feature-based analysis of the natural revocability concept by investigating some features commonly used in signature processing in both the original and newly generated signatures of a group of writers, and exploring the relationship between features, signature style and some demographic factors (specifically, age, gender and handedness) and their effect in relation to the original signatures and new signatures.

Chapter 6: Developing features to improve handwritten signature biometrics

This chapter will explore the development of a type of feature for signature processing which appears to be of particular relevance to the study reported here. The feature relates specifically to the concept of hesitancy (or its converse, fluency) in handwriting, and we will investigate its impact on signature development in the context of natural revocability and signature verification more generally, using an objective measure of the power of the feature. A brief review which covers all the relevant studies and background information about hesitation in different areas will also be given.

Some of the work to be reported in this chapter is published in [203]

Chapter 7: Ancillary issues and final remarks

This chapter will briefly document a number of additional pieces of experimental and analytical work carried out to help to complete a comprehensive picture of related aspects of handwriting biometrics which this study has contributed, as well as a final discussion of all the contributions made in this thesis, together with possible future work required to develop improved strategies for handwritten signature biometric systems in the light of the findings emerging from the reported study.

Chapter 2:

Experimental Infrastructure

This chapter will provide an overview of the basic important tools and infrastructure of all the experimental studies reported in this thesis. Beginning with a brief documentation of the basic signature processing system (utilised in this study) in Section 2.1, Sections 2.2, 2.3 and 2.4 will present the details of data acquisition, feature extraction and some feature processing (respectively) used in the experiments and analysis for this study. Section 2.5 will then present the classification software used for the experiments reported in this thesis. Section 2.6 will discuss the available online signature databases, which the literature shows to have been commonly used previously for signature studies and also the databases and datasets utilised in carrying out the experiments (to be described in greater detail in subsequent chapters) for this study. Finally, Section 2.7 will briefly conclude this chapter. Beginning with this overall consideration of the basic tools and experimental infrastructure will make the subsequent chapters, which report and analyse a range of experimental studies, more cohesive and easier to follow.

2.1 Basic signature processing

Like the basic biometric system, a typical signature processing system is also operated in the following stages: acquisition, pre-processing, feature extraction, matching and a database to store signature templates (as described in Section 1.4.3 in Chapter 1). As described in Chapter 1, at first in the acquisition stage handwritten expression which forms the signature pattern is captured as a two-dimensional signature image (usually associated with offline signature capture) or a set of data packets tracking pen movements in time (usually associated with online signature capture) [17], [204]–[206]. Following the acquisition stage feature extraction is performed to extract the dynamic and static features of the signature, where dynamic features are obtained from the constructional aspects of signatures while static features are obtained from the image of signature [207].

Then, in the pre-processing stage a *feature normalisation* and a *feature correlation* can be performed. In the feature normalisation step, extracted features are normalised, so that the specific feature values measured from each individual signature sample are presented within a fixed range. In feature correlation step, the correlation between the extracted signature features are evaluated, so that redundant features can be removed in order to use more discriminating features for later processes. Sometimes, a *feature selection* is also performed in order to identify and use only the most powerful and highly discriminatory features to obtain the best recognition performance. Finally, in matching stage, *classification* is carried out by comparing the signature features against the stored signature templates in the database, in order to confirm/deny an individual's claimed identity (verification) or to identify the individual (identification). All these steps of a basic signature processing system are shown in Figure 2.1.

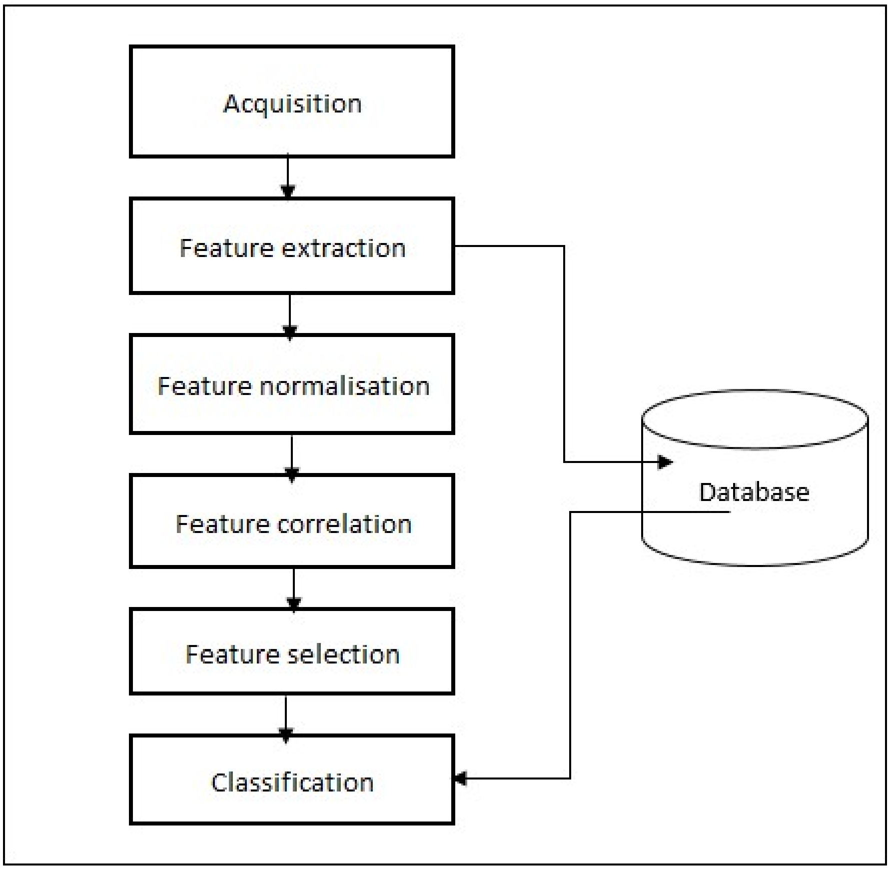


Figure 2.1. Initial steps of the signature processing system used in this study

The basic signature processing system used for the study reported in this thesis implements the processing chain described above and shown in Figure 2.1. The techniques used in each of the steps are described in more detail in the next sections of this chapter.

2.2 Data acquisition

Handwritten signature data were acquired together with some short non-signature handwritten samples for the purpose of the study reported in this thesis. All data were captured using a standard pen of familiar style and feel, and an electronic graphics tablet connected to a computer. The system allowed a subject to write normally on a sheet of paper overlaid on the tablet surface, with the pen movement tracked and a representation stored in the computer in the form of a sequence of time-stamped spatial pen coordinates (e.g. x, y position, pen pressure etc.). The electronic tablet used to capture the data was a Wacom Intuos3 Graphics tablet (model PTZ-930) [146] and capture software developed and refined over a period of time in the School of Engineering and Digital Arts at the University of Kent [208] (called MEDDRAW Data Capture v 2.1) was used for the acquisition and information transmitted to the PC's USB port by the tablet. Further details of the acquisition system are described in Chapter 3.

2.3 Feature extraction

Over the many years of research in the automated processing of handwriting, researchers have used a very large number of both dynamic and static features suitable for signature processing, details of which can be found in the extensive literature which is available (see, for example, [201], [209]–[211]. This makes a choice of which features to work with necessary for any specific study. Since it is not practical to use every possible feature encountered, a total of 60 commonly used features for signature processing [153], [209]–[212] are extracted from the acquired signature samples for the study reported in this thesis. These features are listed and described briefly in Tables 2.1 (features 1 to 20), Table 2.2 (features 21 to 40) and Table 2.3 (features 41 to 60). Although the features listed in the tables are self- explanatory, a further description of these features can be found in [153], [203], [209]–[215].

Table 2.1. Extracted features (1-20)

Feature Type	Feature Number	Feature Names
Static	1	Total distance of pen travelled
Dynamic	2	Total signature execution time
Dynamic	3	Pen lift:(Number of pen ups=> button 1 to 0)
Dynamic	4	Average velocity in X direction
Dynamic	5	Average velocity in Y direction
Dynamic	6	Amount of zero velocity in X direction
Dynamic	7	Amount of zero velocity in Y direction
Dynamic	8	Maximum pen velocity in x - Average pen velocity in x
Dynamic	9	Maximum pen velocity in x - Minimum pen velocity in x
Dynamic	10	Maximum pen velocity in y - Average pen velocity in y
Dynamic	11	Maximum pen velocity in y - Minimum pen velocity in y
Dynamic	12	Maximum pen velocity in x - Minimum pen velocity in y
Dynamic	13	Average pen acceleration in x
Dynamic	14	Average pen acceleration in y
Dynamic	15	Number of zero acceleration sample points in x
Dynamic	16	Number of zero acceleration sample points in Y
Dynamic	17	Maximum pen acceleration in x - Average pen acceleration in x
Dynamic	18	Maximum pen acceleration in x - Minimum pen acceleration in x
Dynamic	19	Maximum pen acceleration in y - Average pen acceleration in y
Dynamic	20	Maximum pen acceleration in y - Minimum pen acceleration in y

Table 2.2. Extracted features (21-40)

Feature Type	Feature Number	Feature Names
Dynamic	21	Maximum pen acceleration in x - Minimum pen acceleration in y
Dynamic	22	Azimuth
Dynamic	23	Altitude
Dynamic	24	Pressure
Static	25	Number of points comprising the image
Static	26	Sum of x coordinate values
Static	27	Standard deviation of x coordinate values
Static	28	Maximum x coordinate value - Last x coordinate value
Static	29	First x coordinate value - minimum x coordinate value
Static	30	Last x coordinate value- minimum x coordinate value
Static	31	Average x coordinate value
Static	32	Maximum x coordinate value - Average x coordinate value
Static	33	Average x coordinate value - minimum x coordinate value
Static	34	Sum of y coordinate values
Static	35	Standard deviation y coordinate values
Static	36	Maximum y coordinate value - Last y coordinate value
Static	37	First y coordinate value - minimum y coordinate value
Static	38	Last y coordinate value- minimum y coordinate value
Static	39	Average y coordinate value
Static	40	Maximum y coordinate value - Average y coordinate value

Table 2.3. Extracted features (41-60)

Feature Type	Feature Number	Feature Names
Static	41	Average y coordinate value - minimum y coordinate value
Static	42	Horizontal centralness
Static	43	Vertical centralness
Static	44	Width of signature
Static	45	Height of signature
Static	46	Width/Height ratio
Static	47	Signature area
Static	48	Width / Area
Static	49	Height / Area
Static	50	Number of vertical midpoint crossing the signature
Dynamic	51	Total time of zero velocity/execution time in X direction
Dynamic	52	Total time of zero velocity / total time in Y direction
Dynamic	53	Average resultant velocity
Dynamic	54	Total time of zero velocity / total time in resultant
Dynamic	55	Amount of zero velocity in resultant
Dynamic	56	Total pen up time
Dynamic	57	Total pen up time/total time
Dynamic	58	Average pen acceleration in resultant
Dynamic	59	Average pen jerk in x
Dynamic	60	Average pen jerk in y

2.4 Feature normalisation and correlation

The *feature normalisation* step is carried out after the features are extracted using a Mean and Variance Normalisation technique (MVN) [153] as defined in (2.1).

$$z_k^i = \frac{x_k^i - \mu_k}{\sigma_k} \tag{2.1}$$

where x_k^i is the k_{th} feature of the i_{th} sample for i = 1, 2, . . ., m, and m being the number of samples and k= 1, 2, . . . , n, and n is the number of features; z_k^i is the corresponding mean and variance normalised k_{th} feature of the i_{th} sample; μ_k and σ_k are the mean and standard deviation (respectively) of all samples in the k_{th} feature as defined in (2.2) and (2.3) respectively.

$$\mu_k = \frac{1}{m}\sum_{i=1}^{m} x_k^i \tag{2.2}$$

$$\sigma_k = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(x_k^i - \mu_k)^2} \tag{2.3}$$

Correlations between all MVN normalised features are evaluated in the *feature correlation* step by using Spearman's rank correlation [216][153]. This is a nonparametric (distribution-free) rank-based estimate of correlation where data are converted to ranks (i.e. ranking all the observation values of a feature from smallest to largest) before calculating the coefficient. The coefficient is calculated between all possible combinations of two feature vectors resulting in F*F coefficient values, where F is the number of features as defined in (2.4).

$$\rho = \frac{\sum(x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum(x_i - \bar{x}_i)^2 \sum(y_i - \bar{y}_i)^2}} \tag{2.4}$$

Where x_i and y_i are vectors of ranks of observations of feature x and y respectively (for $i = 1, 2, \ldots, m$, and m is the number of observations), and $\overline{x_i}$ and $\overline{y_i}$ are the mean of x_i and y_i respectively. Each value of rho ρ (correlation coefficient) obtained by the evaluation is a number between –1 and 1 that determines the extent to which the feature values are related. A ρ value close to zero indicates there is no evidence of any correlation or relationship, while the closer to 1 is this value, the stronger is a positive correlation (i.e. if the value of one feature increases, the feature value in the other feature also increases) while the closer to –1, the stronger is the negative correlation (i.e. if the value of one feature decreases, the value in the other feature increases).

2.5 Classification software

A K-Nearest Neighbour (KNN) classifier [217]–[219] is implemented for the experimental study reported in this thesis using, a data mining package named 'Weka' [220]. The KNN classifier is simple, non-parametric and does not require any explicit training phase. The K-nearest neighbour is determined based on a distance measure appropriate for the extracted feature type (e.g. squared Euclidean distance metric is utilised in this study). Distances are measured between the test signature sample and all the training signature samples representing classes or categories. Then the measured distances are sorted from minimum to maximum and K-nearest neighbours are determined based on K-th minimum distance. The class label appearing the most within the K-nearest neighbours is assigned to the test sample.

2.6 Publicly available online signature databases

The availability of suitable biometric data is a key element for experimentally assessing a systems' performance through benchmarking and evaluation. However, factors such as the nature of the natural variability in handwriting, the occurrence of different types of forgery and additional legal issues regarding data protection, make signature data collection from a large population of individuals for more than one acquisition a time consuming and complicated process. For these reasons the number of publicly available online signature databases is quite limited. Next in this section the details of some public domain signature databases together with some small custom or proprietary databases are briefly outlined.

2.6.1 MCYT

The MCYT database is a bimodal biometric database consisting of online signature and fingerprint modalities [124]. A WACOM Intuos A6 pen tablet was used for signature data acquisition with 100Hz sampling frequency and an overall capture area of 127 mm x 97 mm (width x height) which was further divided into acquisition frames measuring 37.5 mm x 17.5 mm in size. Each target user produced 25 signature samples in groups of 5 and for each user 25 shape based skilled forgeries were also captured from 5 different "impostors". In total, data were collected from 330 users. The signature corpus of the MCYT database was released by the Biometric Recognition Group–ATVS in 2003 [221]. Figure 2.2 shows some signature samples from the MCYT database.

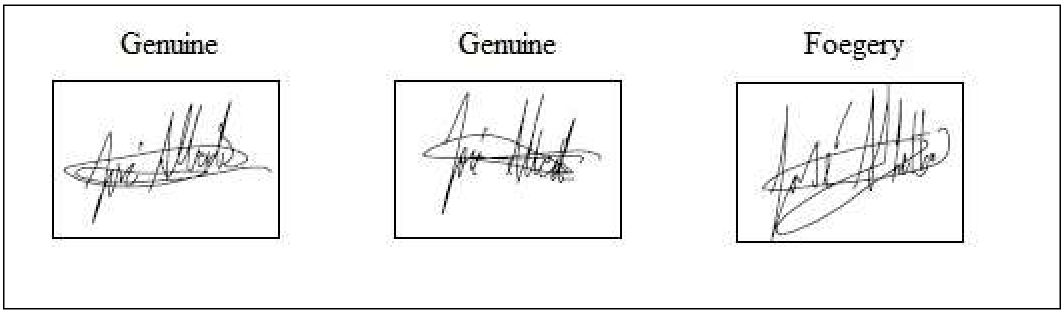


Figure 2.2. MCYT signature samples.

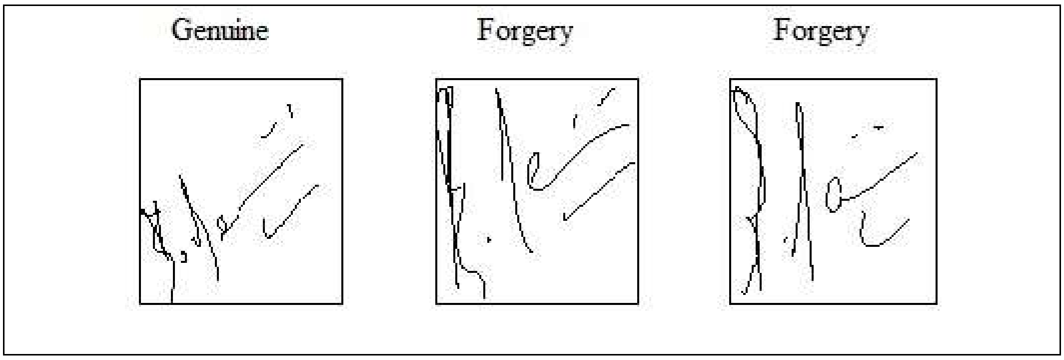


Figure 2.3. BIOMET signature samples

2.6.2 BIOMET

The BIOMET [222] database is a multimodal biometric database consisting of five biometric modalities - speech, face, hand, fingerprint and signature. A WACOM Intuos2 tablet with a sampling frequency of 200 Hz was used for signature acquisition. Data were collected by using both a grip pen that does not provide any visual feedback while writing and an inking pen over a standard paper positioned on the tablet that allows the writer to write conventionally using pen and paper. The grip pen was used in the first session and the inking pen was used in the remaining sessions. In total, signature samples were acquired in three acquisition sessions, with three and five months of interval between them. 15 genuine and 17 forgery signature samples were collected for each user. In total data were collected from

130 users in the first session, 106 users in the second and 91 users in the last session. An example of the signature samples collected in BIOMET database is shown in Figure 2.3.

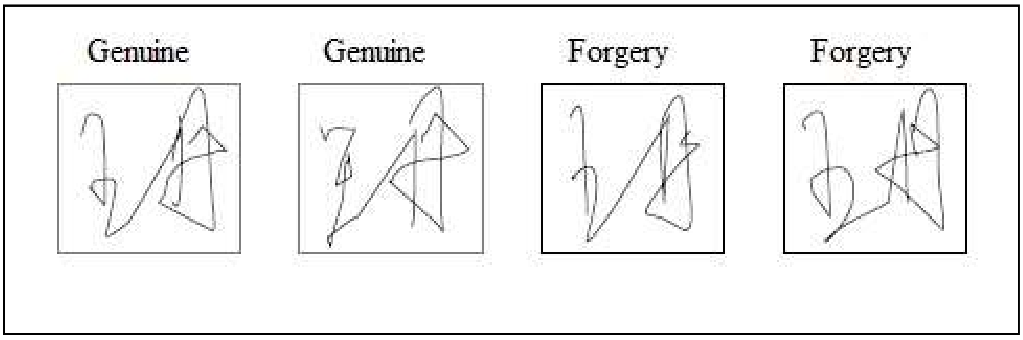


Figure 2.4. SVC signature samples

2.6.3 SVC

The SVC signature database was collected for the First International Signature Verification Competition organised in 2004 to provide a common benchmark for comparing different signature verification systems on the same data and evaluation protocol [223]. Signature data were acquired using the WACOM Intuos tablet with a grip pen (no visual feedback when writing) in two capture sessions with a time period of one week between them. Each target user contributed 20 genuine signature samples (10 in the first session and another 10 in the second session). 20 skilled forgery samples per user were also collected from at least four other users. In order to protect users' personal data, users were advised not to use their original signatures which they use in daily life, instead they used a different signature invented for the purpose of this data acquisition. Some signature samples collected for SVC 2004 database are shown in Figure 2.4.

2.6.4 BioSecure Multimodal Database (BMDB)

The BioSecure Multimodal Database (BMDB) consists of different biometric modalities such as, face, fingerprint, hand, iris, signature, and speech. 11 European institutions participating in the EU-funded BioSecure Network of Excellence [224] were involved in the collection of this database. Data collected in three different sets in three different capture environments: DS1 (internet-based), DS2 (desktop-based) and DS3 (acquisition via a mobile device). Signature samples were collected in DS2 using the WACOM Intuos 3 A6 (at 100 Hz sampling rate) and in DS3 using the HP IPAQ hx2790 PDA (at 100 Hz) in two acquisition sessions [3] In each session 15 genuine and 10 forgery samples were collected from each subject for both sets. In total, therefore, 6300 genuine signature samples were collected from 210 users in the DS2 collection. Some signature samples collected for the BioSecure Multimodal Database (BMDB) is shown in Figure 2.5.

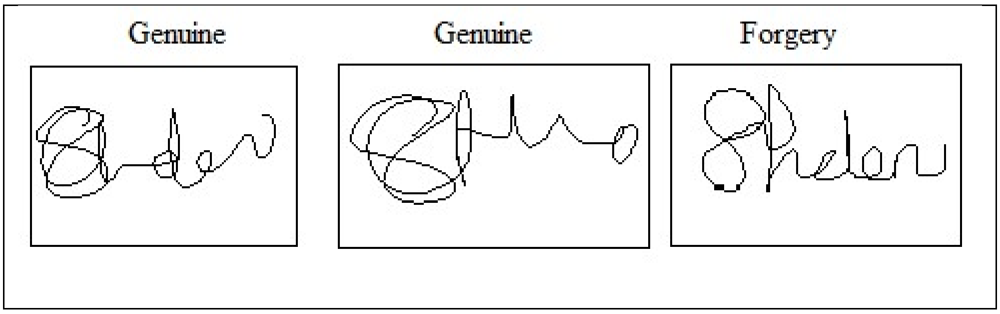


Figure 2.5. BMDB signature samples

2.6.5 BioSecure Kent

The BioSecure Kent database [225] is a multimodal database collected as part of the Europe-wide project undertaken by the BioSecure Network of Excellence [224], [225] in the School of Engineering and Digital Arts at the University of Kent. In fact, this dataset was submitted as a Kent contribution to the wider-ranging BMDB described in Section 2.6.4 above. For signature acquisition of this database, a Wacom Intuos 3 A6 graphics tablet with a sampling frequency of 100 Hz was used. Data were collected in a standard office environment under the guidance of a supervisor [3]. The BioSecure Kent database contains samples from 79 users, collected in two sessions. Each user donated 30 genuine (15 in each session) and 20 skilled forgery samples (10 in each session). Figure 2.6 shows some signature samples collected for this database.

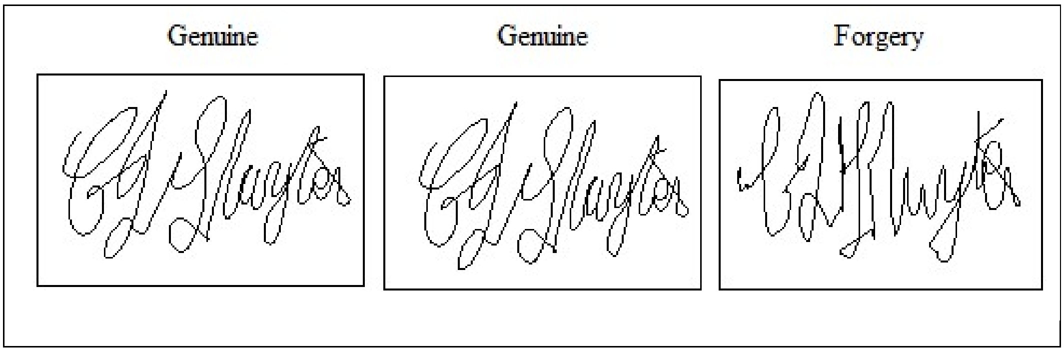


Figure 2.6 BioSecure Kent signature samples

2.6.6 Limitations of currently available signature databases

As described above, due to the nature of the intrinsic variability in handwritten signature and other data collection and protection related issues the number of signature databases readily available to the research community are quite limited. Some of the major and most commonly adopted signature databases are described above. Most of the databases mentioned above contain a good number of subjects' original/genuine signature samples with some forgery samples, providing the opportunity for further analysis and the evaluation of signature processing techniques. But for the investigation of the natural revocability (mentioned in

Chapter 1), which is a major part of the study to be reported in this thesis, two major requirements had to be fulfilled. Firstly, as well as what might be termed a user's "original" signature samples (i.e. his/her currently established signature which s/he uses in her/his everyday life) samples of a newly developed signature, such as might be adopted should the original need to be withdrawn, needed to be collected from the same subject. In the SVC database, newly invented signature samples were collected in two sessions, which were classed as genuine signature samples (for not making the actual signature publicly available which the participant uses in daily life), but the database does not have the actual original signature samples while other databases have the original signature samples as "genuine" samples but there is no newly invented signature samples. Secondly, signature samples needed to be collected over a longer period of time (e.g. four sessions or more with a period of at least a week between them) using the same acquisition device from the same subjects to study how increased familiarity with a new signature affects its form and reproducibility. In most of the databases described above signature data were collected in only one or two capture sessions. In the BIOMET database, signature samples were collected in three capture sessions but using two different types of tablet pens. Unfortunately, the publicly available databases do not facilitate the experimentation proposed. All these deficiencies in relation to the requirements of the proposed study make the compilation of a new and "bespoke" database a fundamental part of the work in the study of interest here. This new data collection exercise will produce a database of both original and newly invented handwritten signature samples collected over a considerably longer period of time (maximum of ten sessions) when compared with the acquisition process of other databases currently available in signature biometrics.

Details of acquisition process, the collection protocol, and the contents of this new database, which we have called the 'Revocability database', are described in detail in Chapter 3, and this database forms a significant contribution of the overall study undertaken. As well as this Revocability database, the BioSecure Kent database has also been used for the experimental analysis reported in this thesis. For ease of subsequent reference, the Revocability database and the BioSecure Kent database are designated as the Rev-Kent and Bio-Kent databases respectively.

2.7 Conclusion

In this chapter, the basic experimental framework and important practical details used in the experiments and analysis (to be described) have been introduced.

Initially the steps of a basic signature processing system and a basic signature processing system were briefly described. Taking this as a basis for the signature processing system utilised for the experiments reported in this thesis, the techniques used for data acquisition, feature extraction, feature normalisation, feature correlation, classification were presented.

A review of some of the principal online signature databases available to researchers has been presented, providing a useful and critical analysis of the potential additional benefits and enhanced characteristics of the proposed new data acquisition exercise. Finally, the two databases utilised for the experiments carried out were identified.

The next chapter will present the details of the new data acquisition exercise.

Chapter 3:

Revocability Database Compilation

This chapter will discuss the handwritten signature data collection protocol together with an overview of the acquisition system. Section 3.1 will give a brief background of the means and methods used for the construction of a handwriting and handwritten signature database. Section 3.2 will present the data collection protocol and ethical approval procedure. Section 3.3 will discuss the data acquisition system – hardware and software. Section 3.4 will illustrate the collected data and subject information acquired. Section 3.5 will discuss the challenges faced during data collection and explain the value of this database and Section 3.5 will conclude the chapter.

3.1 Introduction

The overall aim of the data collection procedure is to establish a database of handwritten samples based on the handwritten signature, but enhanced by the addition of short, simple non-signature handwriting samples. Most importantly, in order to enable the investigation of "natural revocability", as mentioned in Chapter 1 (detail study of the natural revocability will be presented in Chapter 4), it is necessary to acquire samples of a new signature developed by each participant to simulate a situation where an original signature has been compromised, and a new one is required. Moreover, it is necessary to obtain samples over a period of time, to study how increased familiarity with a new signature affects its form and reproducibility. As discussed in previous chapters, when considering sample capture for automated processing, signature data are traditionally divided into two categories: online – where both the spatial and temporal information regarding the signature (i.e. information about both the appearance and execution of the signature) is available from the written input; and offline –where temporal information is not available, but only the spatial information (i.e. form and appearance), typically from a scanned document is available [17], [119], [198].

Signature data captured through offline methods may involve the use of a scanner or a camera as input device and therefore only the 2D signature image is captured. On the other hand, the more common means of capture, that of online signature acquisition, typically uses a graphics tablet device or digitiser, although the use of a camera based acquisition system, incorporating a visual tracker of the pen-tip position in the writing surface, has been proposed in the literature [226], [227]. The use of digitising tablets or instrumented pens goes back as early as late 1970s according to data input devices reviewed by Plamondon and Lorette [228]. The types of error (spatial errors, temporal errors and intrinsic errors) that are likely to occur during data collection using digitisers were reported by Meeks and Kuklinski [229]. Digitiser technology has since developed further providing high accuracy and reliability, and user- friendly interfaces while introducing the measurement of parameters in addition to the standard x y coordinates, with higher sensitivity in the pen tip pressure captured, and other characteristics such as pen altitude, azimuth, etc. A large number of descriptive parameters can subsequently be extracted from those directly captured by the tablet, such as pen tip velocity, acceleration, and so on.

The next sections of this chapter will describe the digitizer tablet used in this study, and the data acquisition software used to collect the data along with the data collection protocols.

3.2 Data collection protocol

To facilitate the data collection, an appropriate and robust data collection protocol was required in order to guide and underpin the construction of a viable and potentially extremely valuable and unique database of handwritten signature and non-signature handwriting samples, incorporating samples illustrating the effects of "inventing" a new signature form. The data collection protocol describes the strategy adopted to ensure the uniform collection of the handwritten signature and non-signature handwriting samples, defining systematic and reliable procedures to ensure high quality handwritten data was collected under precisely defined and repeatable conditions. The protocol also provided the ability to manage and administer the data acquisition process in a systematic way and uniform. This specified the overall procedure for the acquisition of the signature and non-signature handwriting samples, covering important issues such as the following:

- What equipment will be used for acquisition: All signature and non-signature handwriting samples will be captured using a standard pen of familiar style and feel, and an electronic graphics tablet connected to a computer. The system allows a subject to write normally on a sheet of paper overlaid on the tablet surface, with the pen movement tracked and a representation stored in the computer in the form of a sequence of time-stamped spatial pen coordinates. Details of the acquisition system are described in Section 3.3.

- Task division within the acquisition process: There are three parts of this acquisition process: Part A, where a subject will be asked to provide samples of his/her usual signature; Part B where a subject will be asked to provide samples of a new signature invented by the subject or suggested by the researcher, for the purpose of improving the understanding of the notion of 'natural revocability' for the handwritten signature; and Part C, where samples of handwritten examples of numerals and alphabetic strings will be collected.

- Number of acquisition sessions: Volunteers may be asked to return and repeat some or all of the data collection process, both to increase the number of available samples per user, but, more importantly, also to reflect changes in handwriting styles and appearance with time. Not all subjects will take part in all parts of the collection process.

- Recruitment: A varied population is to be recruited with respect to age, gender, and so on. A private space will be provided for the data collection, and a supervisor will be present throughout every collection session.

- Data storing: The samples collected will be stored so that they are linked to a reference number rather than the volunteer's name, and only the research team will be able to link the samples to the volunteer personally. This information will be kept strictly confidential within the research team.

- How the data will be used: The data collected will be analysed by researchers at the University of Kent for the purposes of research into writing and signing, and the development and evaluation of automatic handwriting processing systems and related technologies. The data will not be made available to any third parties. The results of the evaluation will be documented and are likely to be published in the scientific literature to help others benefit from the evaluations in the future, but subject anonymity will always be preserved.

3.2.1 Ethical requirements

A Data collection at the University of Kent involving human participation is subject to approval by the University Ethics Committee, according to rules governing most respectable institutions to ensure the legal protection, privacy and safety of users and their biometric samples, and overseeing this process is part of the function of the Ethics Committee.

The University of Kent Ethics Approval Procedure requires applicants to complete an "Application Form for Ethical Approval from Research Ethics Group", together with the provision of supporting documentation which included the associated participant information sheet, consent form and participant detail sheet, cover letter and an Ethics Review Checklist, all of which are provided in Appendix A.1, A.2, A.3, A.4, A.5, A.6. All of these documents are submitted to the Faculty of Science Ethics Committee where the study is reviewed and any further information and questions from the reviewers can be asked regarding the submitted application. Once the application passes the ethical review procedure, the study can then commence making sure the ethical procedures are adhered to.

3.2.2 Participant recruitment process

After approval from the Ethics Committee the next part of the database compilation process involved recruiting subjects to take part in the data collection. A large number of staff and students from the University of Kent were invited by email together with the member of the general public through personal contact with respect to age, gender, ethnicity, and so on.

A list was compiled of people who agreed to take part in the data collection. They were then contacted by email with an appointment schedule. After confirmation of the appointment date and time they were sent the Participants Information Sheet via email in which they were informed of the purpose of this study, how long this will take, what will happen to the samples provided, how to withdraw from participation, what will happen to the results of the evaluations using the database, information about the research and contact details for further information.

3.2.3 Meeting and greeting participants

The next phase after recruitment involved meeting and greeting participants. Each participant is met by the researcher on arrival, and taken to a private room (to avoid distractions) where the collection infrastructure is set up. The participants were asked if they read and understood the Participant Information Sheet and invited to ask any further queries. Each participant was asked to provide their contact information such as first name, surname, e-mail, and telephone number etc. in the first section of the Participant Details form. An identification number and the date of data acquisition was also recorded in this section. In the second section participants provided anonymised information relating to age, gender, ethnic origin, handedness, occupation.

The participants were also asked to read and sign a consent form, which was used to describe the general purpose of the research project. The information of the consent form is similar to the information of the Participant Information Sheet which listed the following details:

- Data to be collected
- Criteria to be satisfied for participation in the data collection
- Purpose of this collection
- Privacy and safety of participants' samples
- Proposed use of data after collection
- Contact details of the investigators
- Details regarding the opportunity to withdraw participation from this study at any time.

3.2.4 Start of the data collection process

After all the initial paperwork has been completed (described above) and the participant is satisfied with the explanation of the data collection study each participant she/he is asked to make her/ himself comfortable with the writing arrangements – using a standard pen of familiar style and feel, and an electronic graphics tablet connected to a computer, as described in the data collection protocol in Section 3.2. A Details of the acquisition system are described later in this chapter (Section 3.3). Once the participant is comfortable with all arrangements, the data collection process itself can begin. As described in Section 3.2 the data collection process consists of 3 parts -Part A, Part B and Part C and the acquisition starts from Part A.

Figure 3.1 illustrates the introduction to the data collection process.

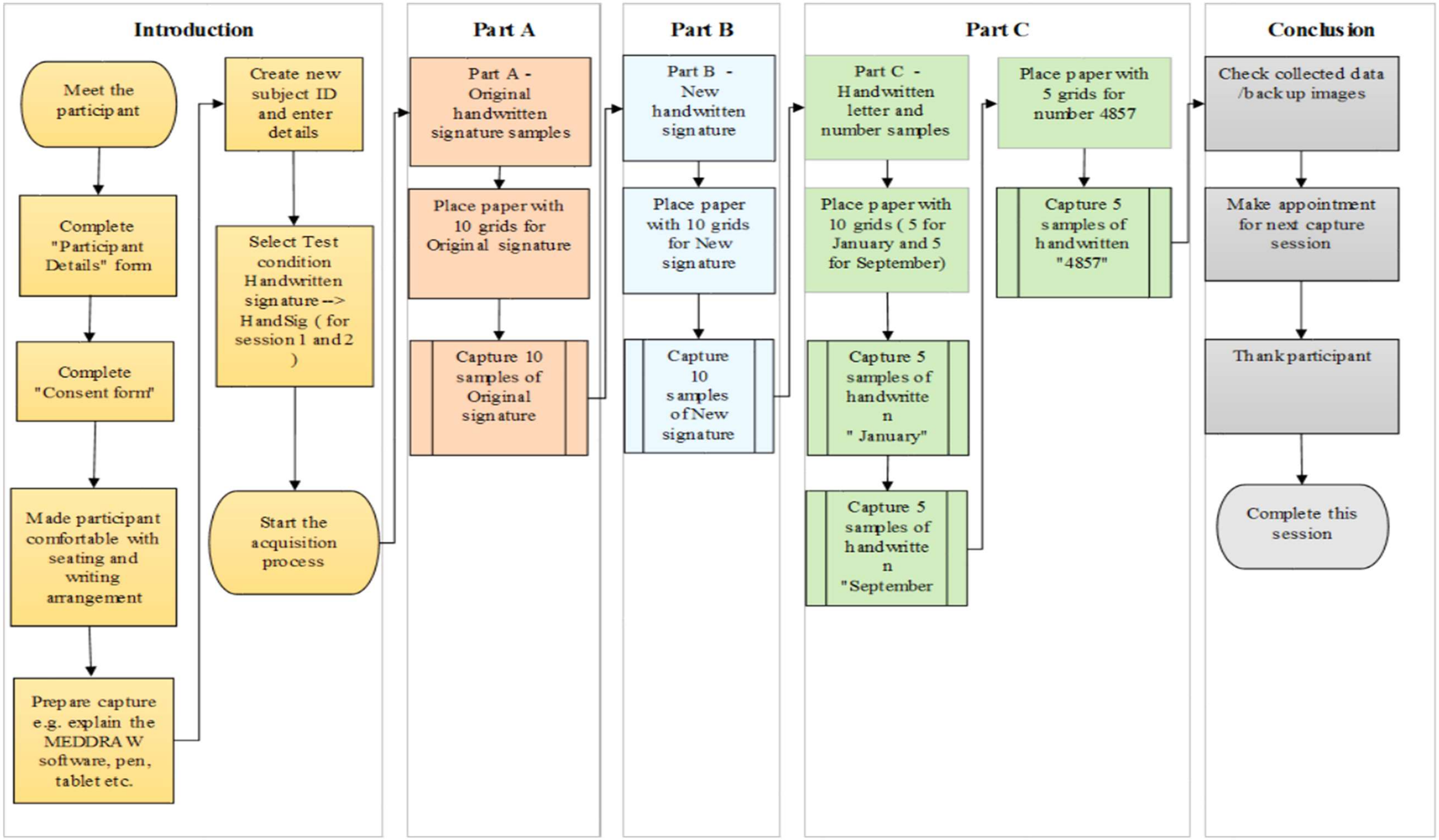
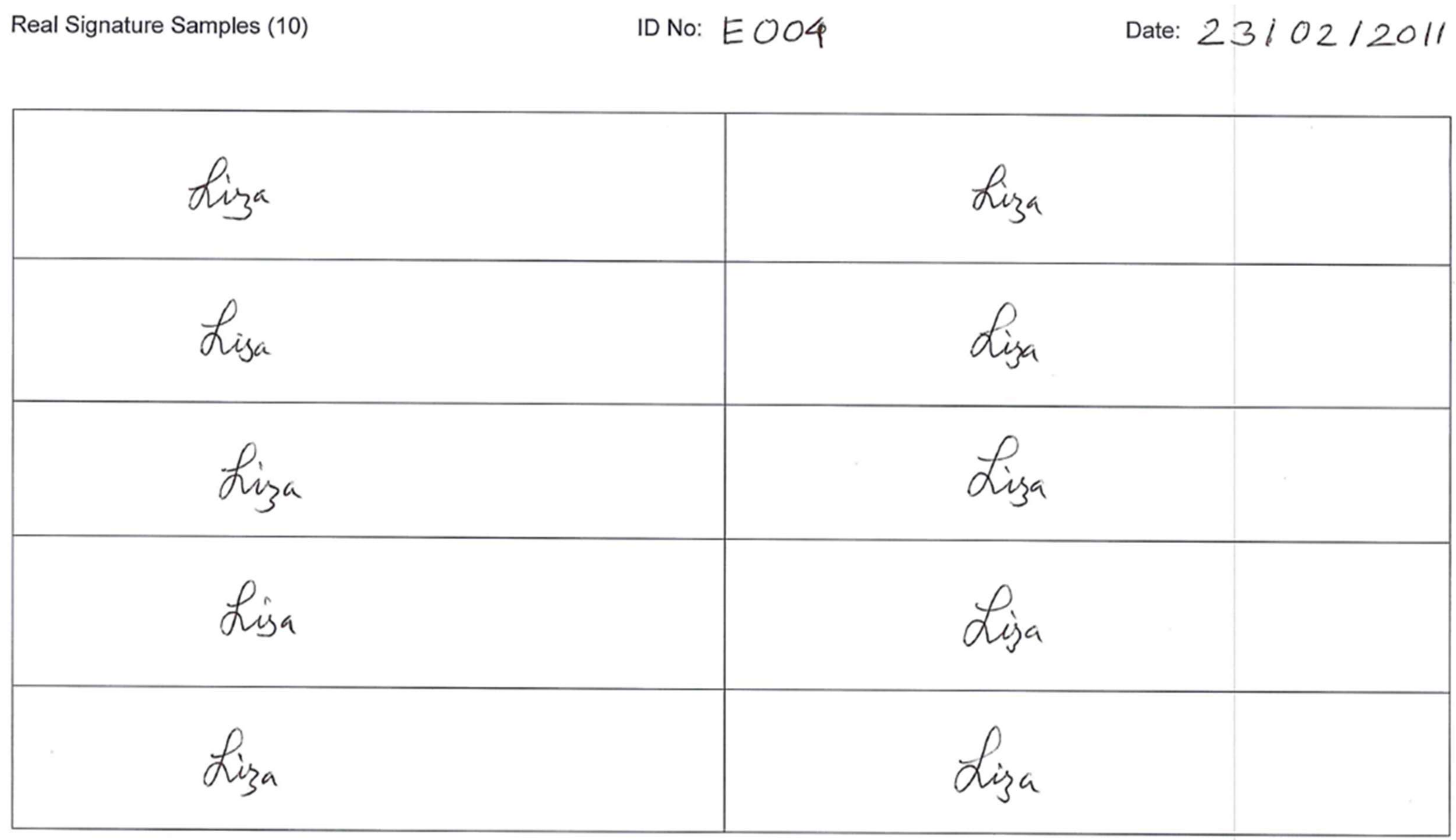


Figure 3.1 Data collection process (First session)



Real Signature Samples (10) ID No: E004 Date: 23/02/2011

Liza	Liza
Lisa	Liza
Liza	Liza
Lisa	Liza
Liza	Liza

Figure 3.2. Part A – Original signature acquisition form

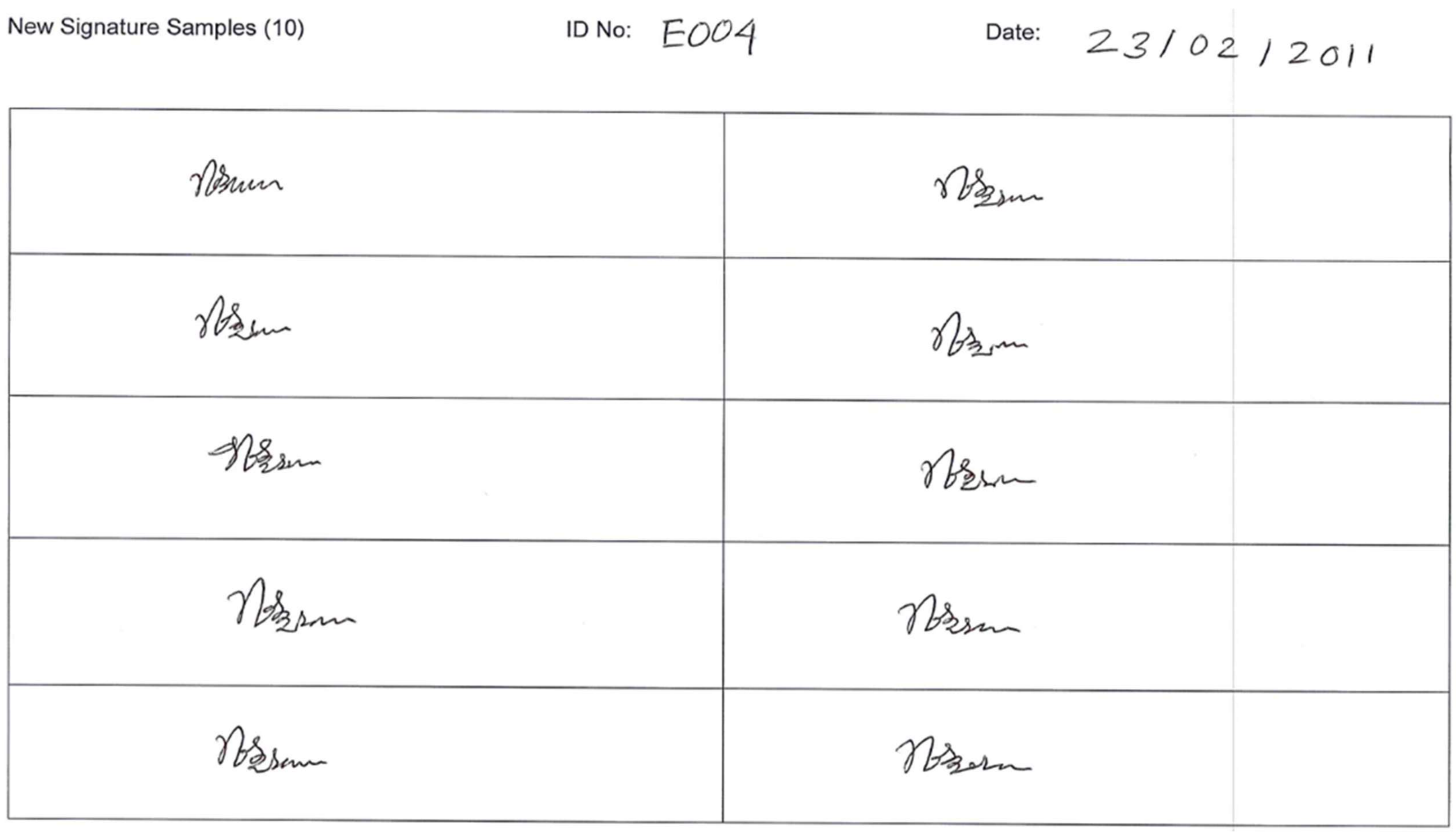
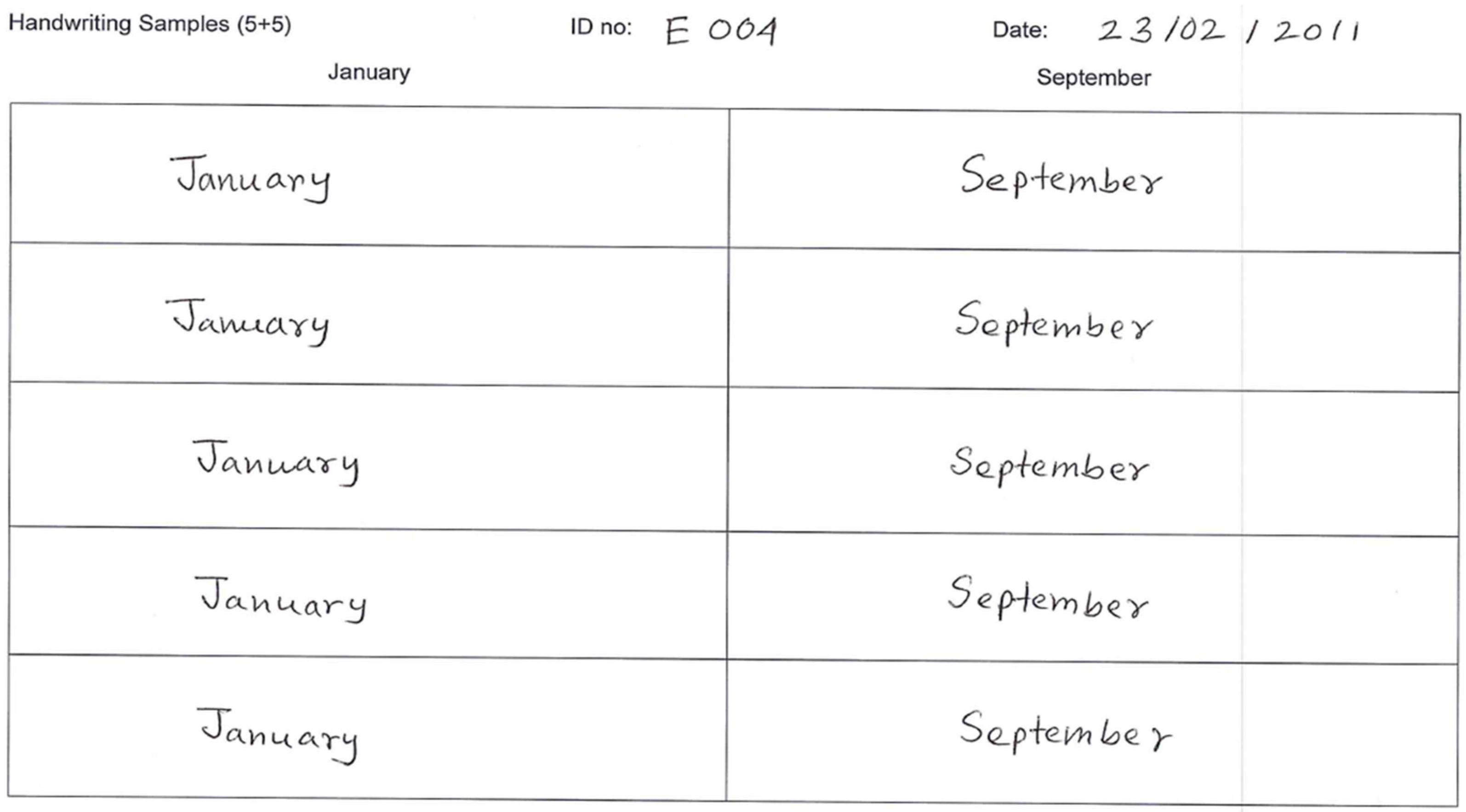

New Signature Samples (10) ID No: E004 Date: 23/02/2011

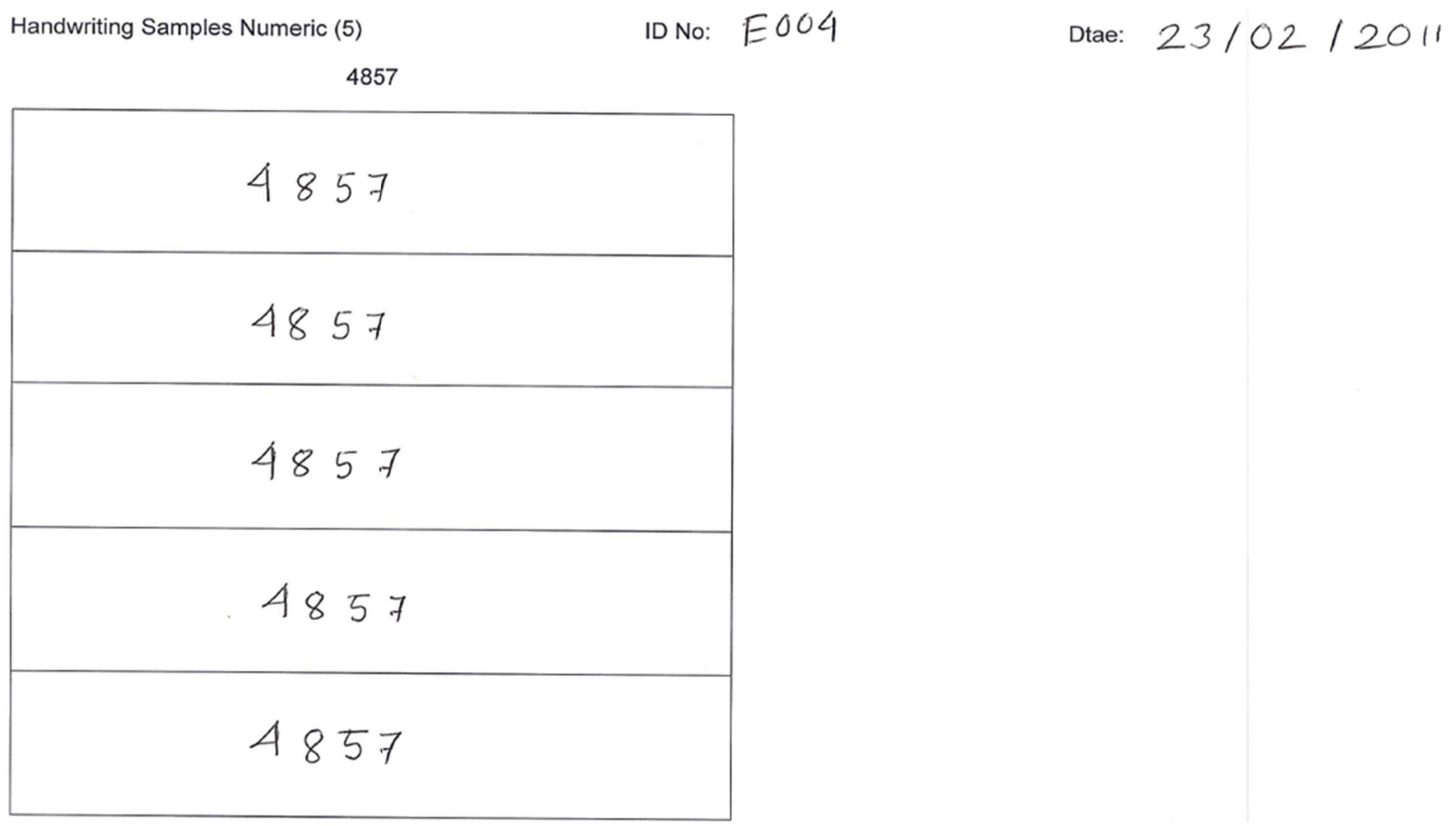
Figure 3.3. Part B – New signature acquisition form



Handwriting Samples (5+5) ID no: E 004 Date: 23/02 / 2011

January	September
January	September
January	September
January	September
January	September
January	September

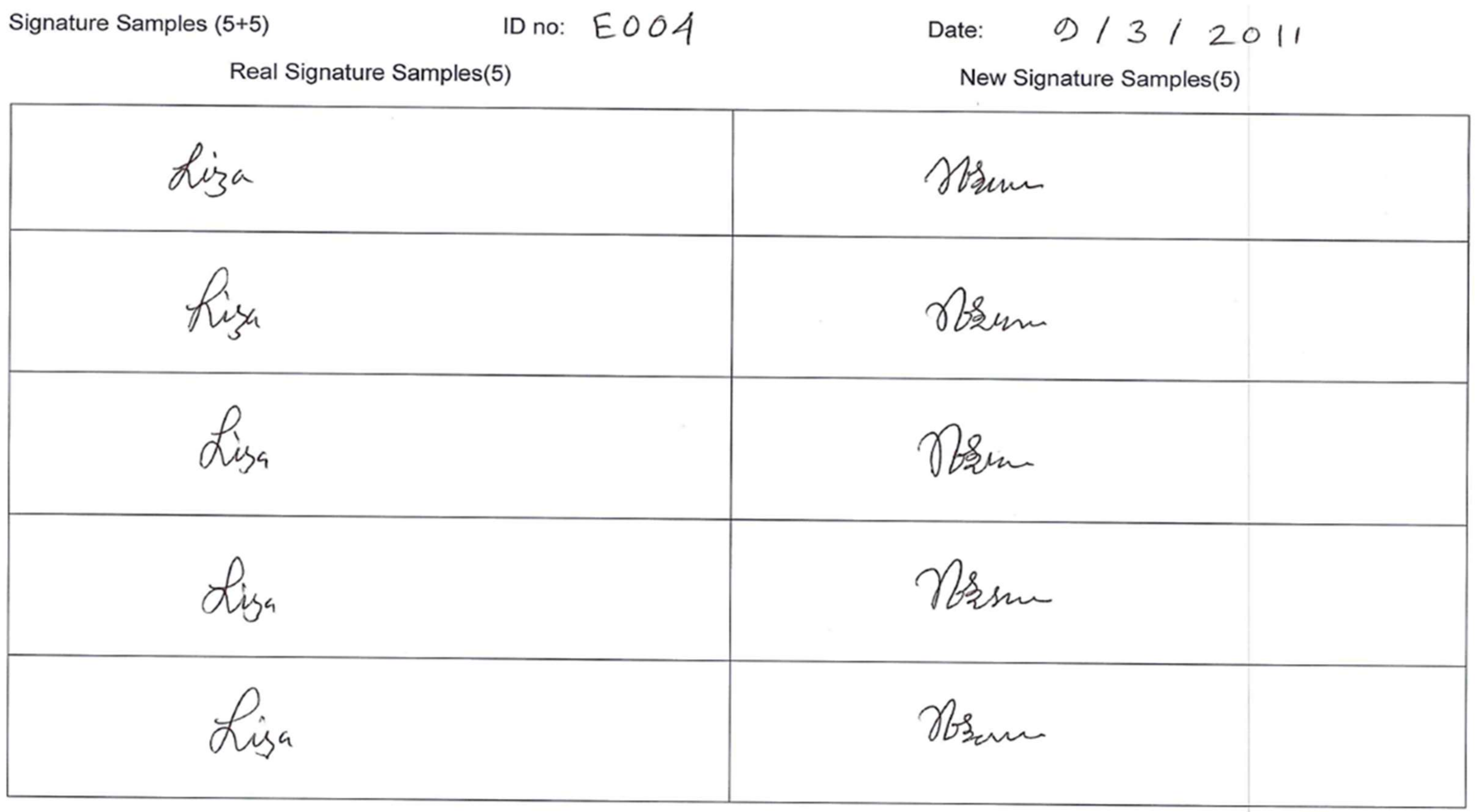
Figure 3.4. Part C– Handwritten words acquisition form


Handwriting Samples Numeric (5) ID No: E004 Dtae: 23/02/2011

4857

4857
4857
4857
4857
4857

Figure 3.5. Part C– Handwritten numerals acquisition form


Signature Samples (5+5) ID no: E004 Date: 0 / 3 / 2011

Real Signature Samples(5) New Signature Samples(5)

Figure 3.6. Part A and Part B signature acquisition form for session 3

3.2.5 Part A acquisition

Under supervised conditions the participant was asked to provide samples of her/his original signature (the signature (s) he use in her/his everyday life) on the sheet of paper provided , one sample at a time in each grid using the pen provided mentioned in 3.2. In session one and two participants were asked to provide ten samples, and, if repeated, five samples from session three onwards. A sample of the Part A capture form is shown in Figure 3.2.

3.2.6 Part B acquisition

Once the Part A acquisition is completed the participant is asked to provide samples of a new signature invented by the participant (using the same acquisition system as in Part A) or suggested by the researcher for the purposes of improving an understanding of how a signing style develops. Blank papers were provided for the participant to try and practice the new signature before writing it on the tablet. Like Part A acquisition, participants were also asked to provide ten samples of the new signature for session one and two, and five samples of the same from session three onwards, if repeated. A sample of the capture form used in the experiment is shown in Figure 3.3.

3.2.7 Part C acquisition

After Part A and Part B each participant was asked to provide samples of handwritten examples of the numerals "0" to "9", and the alphabetic strings representing the months of the year, "January", "September" (using the same acquisition system as in Part A and Part B). A four-digit number '4857' (like a pin number) was chosen to represent the different orientations of numeral (e.g. numerals with horizontal and/or vertical line, angle, curve etc.) as much as possible within the four-digit number. The two months "January" and "September" as alphabetic strings were also chosen to represent as many as different letters possible within the twelve months of the year. The purpose of the Part C collection was to have the database enhanced with signature and non-signature handwriting data from the same participant. As the main focus of this study was to study the behaviour of handwritten signature, collected data in this part (Part C) has not been used for experimental purpose in this thesis. But these data can be used for future analysis such as, combining handwritten numerals or handwritten alphabet strings with handwritten signature for user identification, verification etc. A sample of the Part C collection form is shown in Figure 3.4 and Figure 3.5.

3.2.8 Repeated sessions and conclusion of the data collection

As mentioned in the protocol in Section 3.2 some volunteers were recruited to take part in the repeated sessions with an interval of one week between sessions to observe and reflect changes in handwriting styles and appearance with time. Sessions were repeated from two to ten sessions. In session one and session two all collection phases (Part A, Part B, Part C) were involved, where in Part A ten samples of a subject's original signatures, in Part B ten samples of the new signature and in Part C five samples of writing "January", five samples of handwriting "September" and five samples of handwritten numerals " 4857" were captured. As mentioned earlier, the repeated sessions aim to allow the observation of the changes in handwriting styles and appearance of signatures and also to understand the possibility of exploiting the 'natural revocability' phenomenon in the handwritten signature. A week difference between collecting sessions allowed to observe how the participants naturally adopt the new signing process rather than doing it every day or a long two or three weeks separation and collecting up to ten sessions allowed to observe the variation over a period of time within the bounds of feasibility of this data collection (challenges of this data collection is described later in this chapter in Section 3.5). Samples were collected only from the Part A and Part B phases, five samples in each phase from session three onwards. (An example is shown in Figure 3.6). After the first two sessions participants found using the tablet and pen very straightforward, so the number of samples actually collected was reduced down to five from session three. Also, in other known databases, four or five signature samples were collected per session or per set. [3], [124], [230].

At the end of each session an inspection was made of the dataset to determine if all samples were collected. The participant was thanked for volunteering to take part in data collection study and a date was arranged for the next session where applicable.

3.3 Acquisition System

To acquire the handwriting and handwritten signature samples as outlined in the protocol a digitising tablet and a custom piece of software was required. The hardware and software employed in this data collection are described in the following sections. Table 3.1 shows the general features of the data acquisition process and the biometric data for each subject in each session.

3.3.1 Digitising Tablet

The digitiser used in this data collection exercise was a Wacom Intuos3 Graphics tablet (model PTZ-930). It measured 439.5 mm by 340 mm in dimension with an active area of 304.8 mm by 228.6 mm. It also has a cordless Grip pen, ink pen and mouse which do not need batteries. The pen enables 1024 pressure levels and 5080 dpi resolution to be achieved [231]. The Intuos 3 pen was used because of its capability to be easily used by a wide range of people. For example, it can be used by those who suffer from Repetitive Strain Injury (RSI) as it has a non-slip rubberised grip and is excellent for hand balance. The pen tracks excellently well, introduces absolutely no delay with its movement and complex intercepts. According to the Wacom specification [232] the communication between the tablet and the pen is achieved by means of an electromagnetic process. The tablet transmits an electromagnetic signal to the pen, which in turn modifies it and sends it back to the tablet for position and pressure analysis. A grid of wires below the tablet's screen alternates approximately 20 microseconds to accommodate the transmission and reception of the information. The tablet sends the information to the computer via its USB port. An illustration of the transmission is shown in Figure 3.7.

Table 3.1. General features of the data acquisition process

Subjects	62
Sessions	1-10
Supervisor	Yes
Condition	Standard office (desktop) with indoor light.
Hardware	PC, Wacom tablet

	Session 1- 2		Session 3 – 10	
Signature	**20 samples** 10 original 10 new	**20 .tst files** **20 image files**	**10 samples** 5 original 5 new	**10 .tst files** **10 image files**
Handwriting	**15 samples** 10 words (5 January, 5 September) 5 Numerals (4857)	**15 .tst files** **15 image files**		

In order for the signing process to be as natural as possible it was deemed necessary for the signatures to be executed on paper which is placed as an overlay on the surface of the tablet, while making use of the Intuos 3 Ink tip rather than the polyacetal tip that would allow 'invisible' drawings on the bare surface of the tablet. This ensured that the signing process was entirely familiar and natural for the subject, giving the accustomed visual feedback during signing. As a guide, the A4 paper was divided into five sections (or boxes) both on the left and on the right (as shown in Figure 3.2, 3.3, 3.4, 3.5 and 3.6) and the writing and signing area was restricted to the box provided, as closely simulating restrictions applied in the signing space available in most implementations encountered in common point–of-sales applications.

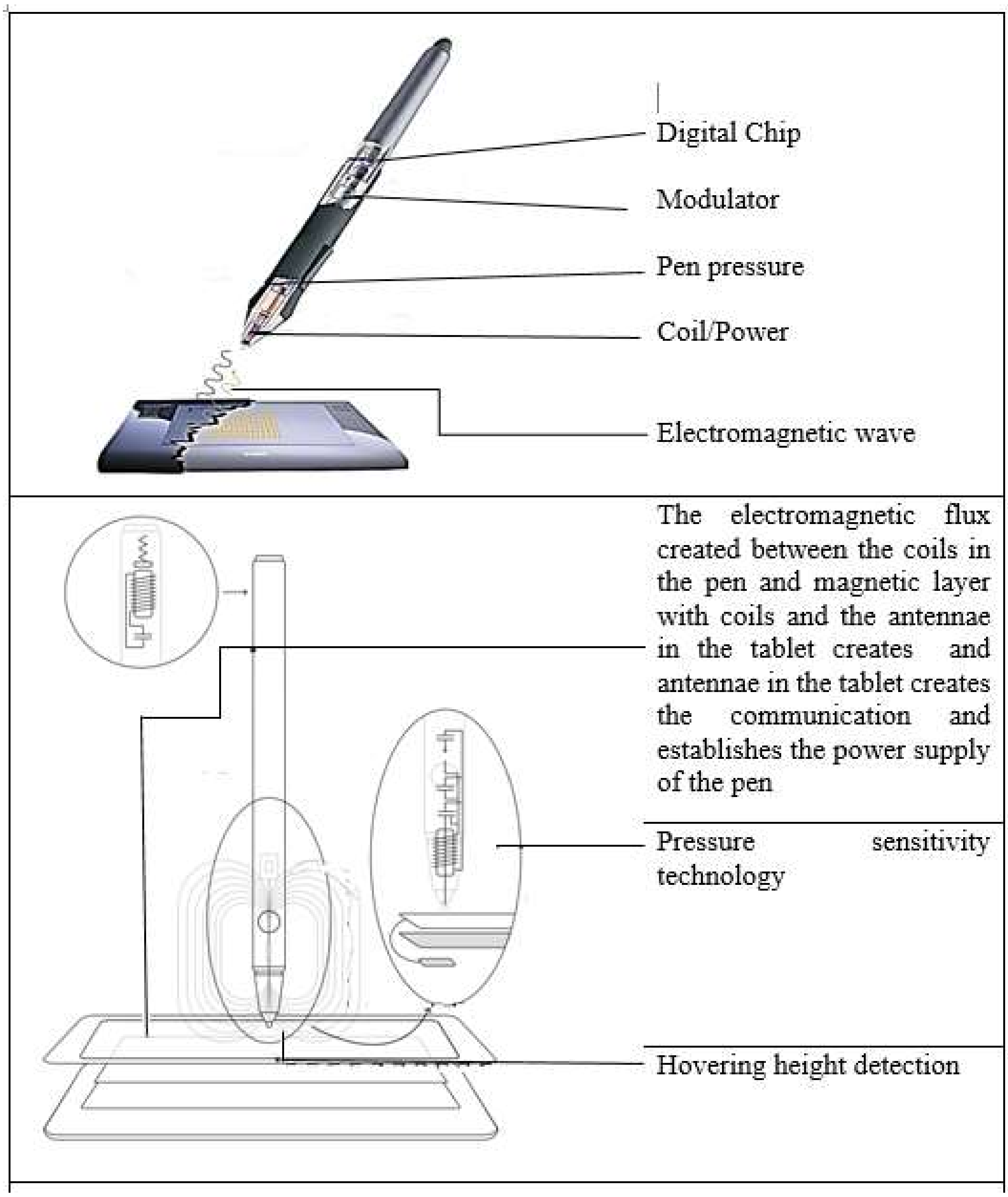


Figure 3.7. Transmission of information between pen and tablet

3.3.2 Acquisition Software

A piece of capture software developed in the School of Engineering and Digital Arts at the University of Kent (called MEDDRAW Data Capture v 2.1) was used for the acquisition and storage of the information transmitted to the PC's USB port by the tablet. It was developed as part of the MEDDRAW project to enable computer-based assessment of hand-drawing tasks to enhance diagnosis and assessment of neuropsychological conditions. But the techniques developed are applicable to any condition where assessment can be aided by objectively studying the patient's performance in writing or drawing tasks such as the copying of visual shapes.

The MEDDRAW software application was developed based on the programming interface for using digitising tablets specified in the 'Wintab Specification' [232]. This software enables the recording of pen movement data from an attached Wintab compliant graphics tablet or a Tablet PC. It was developed to capture multiple attempts (writing or drawing attempts) from an individual test subject and to support multiple conditions and tests. The *condition profile* describes the fields that are to be recorded in the subject info file and all capture (TST is described in Section 3.3.4) files and the *test profile* describes which overlays are to be displayed for the test being conducted. It also describes if the test is to be conducted in portrait or landscape mode, and if a timer is required to sound after a specified interval, for each overlay. Each test attempt is stored in an individual text file (TST) with a range of Wintab parameters including position, pressure, and button status for future analysis. Each data packet is timestamped to microsecond accuracy (actual sample rate dependant on hardware configuration. The details of the parameters are described in Section 3.3.4.

The overlay templates provided in the software were not suitable for this data collection, so it was configured by adding new overlays, a new *condition profile* named “Handwritten signatures” which was segmented into two *test profiles* – ‘*HandSig’* and *‘HandSig3’*, the initial test profile being for the first two enrolments and the latter for the third to tenth enrolments, to display the overlays while capturing data. Figure 3.8 shows the new condition profile ‘Handwritten signatures’ and its two test profiles.

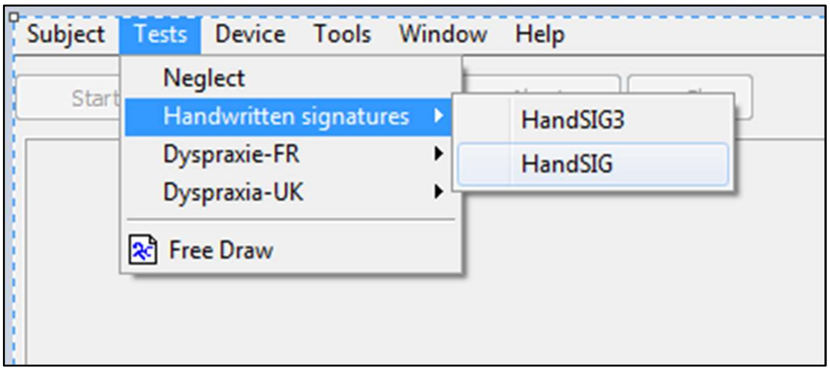


Figure 3.8. Condition and test profile in MEDDRAW

The MEDDRAW software has been designed to allow the incorporation of a tablet PC, a graphics tablet or an oversize tablet. For the purpose of this study as described above, a graphics tablet was used. This mode scales the window of the graphics tablet to the size of the image as seen on the screen. The subject was then selected; this was either a new subject or an existing one. If a new one, a form is displayed which contains vital information of the subject to be inputted. If an existing subject, the capturing mode is turned on. The ID no, name, date of birth, age, gender, nationality, file creation date, file last edit date, writing hand (left or right), session dates and comments are recorded and stored in the database. For a new subject and existing subject, supervision had to be incorporated into capturing the details as described in the data collection protocol.

3.3.3 Data storage

The data captured were stored and written to files and folders in the directory where the program was installed in individual subject directories. As shown in Figure 3.9, each subject directory e.g. E060, contains a ‘subject.txt’ file which is written when the subject is created and contains all of the information entered when the subject was registered, as well as a directory for each capture session. These capture session directories are named by *<Test_Name>-dd-mm-yyyy-mm-ss*, where *dd-mm-yyyy-mm-ss* is the date and time at the start of that capture session. An example directory might be *HandSIG-11-03-2012-19-51-03.* This allows for multiple capture sessions any length of time apart without the previous data being overwritten.

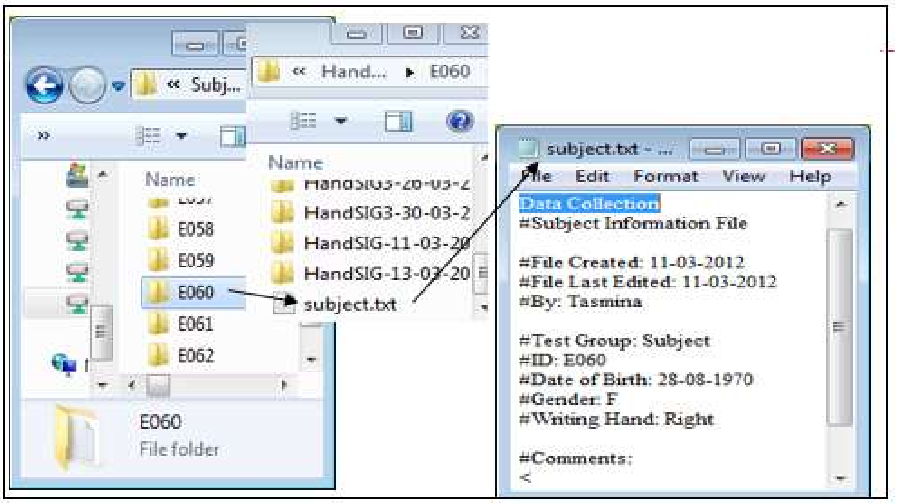


Figure 3.9. Text file containing subject information (for user E060)

3.3.4 TST File Contents

Each session directory of individual subject directories contains a tst file and a jpeg image file for each sample (as shown in Figure 3.10). The tst files themselves contain information about the subject (taken from the relevant subject.txt file at the time of capture), the tablet information (technical data retrieved from the capture device) and finally the data packets themselves, consisting of the following tab separated parameters.

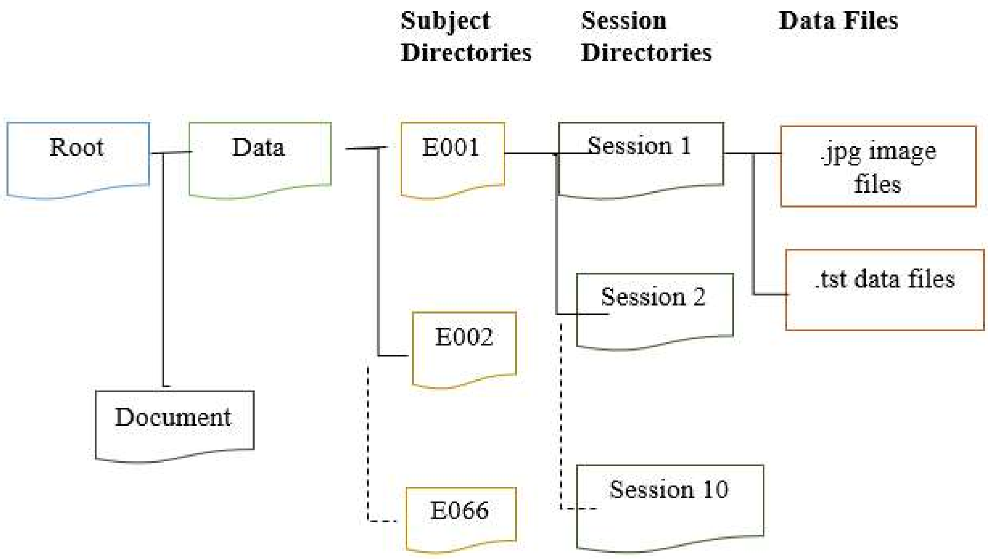


Figure 3.10. Directory structure of the database

1) Timestamp – The offset in milliseconds since the start of the capture process.

2) X coordinate – Horizontal location of the pen.

3) Y coordinate – Vertical location of the pen.

4) Normal Pressure – Normal pen tip pressure.

5) Tangential Pressure – Tangential or barrel pressure.

6) Status – If the cursor is in or out of the context etc.

7) Cursor - The cursor type that generated the packet such as a stylus, or a finger touching a touch pad.

8) Context – The id of the context that generated the packet.

9) Buttons – The button or pen tip state (such as button released/pressed or pen up/down)

Three data items relating to the orientation of the pen.

10) Azimuth - the clockwise rotation of the cursor about the z axis through a full circular range.

11) Altitude - the angle with the x-y plane through a signed, semi-circular range. Positive values specify an angle upward toward the positive z axis; negative values specify an angle downward toward the negative z axis.

12) Twist - the clockwise rotation of the cursor about its own major axis.

Three data items relating to the rotation of the pen.

13) Pitch - The pitch of the cursor.

14) Roll - The roll of the cursor.

15) Yaw - The yaw of the cursor.

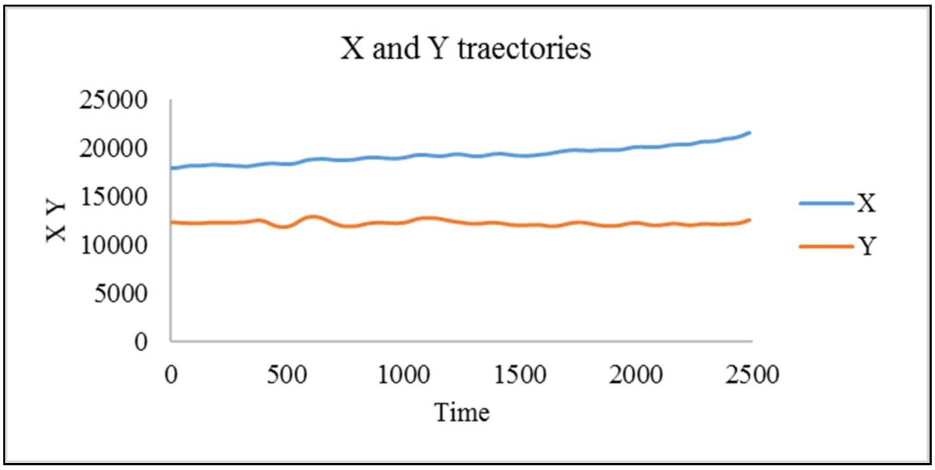
An example of the .tst file contents is shown in Figure 3.11.

```
#Tablet Info
#Protocol: WACOM Tablet
#Capture Device: Graphics Tablet
#Freq: 100
#Axes Unit: 1000*cm
#X_Origin: 0
#Y_Origin: 0
#X_Extent: 30480
#Y_Extent: 22860
#X_Res: 1
#Y_Res: 1
#NP_Max: 1023

#Subject Info
#ID: E004
#Age: 32 years 2 months
#Gender: F
#Writing Hand: Right

#Comments:
<
No Comment
>

#Ti    X      Y      NP   TP   St   Cu   Co    Bu   Az    Al    Tw   Pi   Ro   Ya
#Data:
0      17909  11851  0    0    0    7    517   0    1680  300   0    0    0    0
5      17909  11851  0    0    0    7    517   0    1680  300   0    0    0    0
15     17899  11869  0    0    0    7    517   0    1680  300   0    0    0    0
```

A header at the top shows the mnemonic for each item (NP = Normal Pressure)

Figure 3.11. Example of a .tst file contents

A set of features can be extracted from these captured raw data from the tablet as described in Chapter 2 and utilised for the experimental work reported in later chapters in the thesis. The commonly used raw data items (Timestamp, X Coordinate, Y Coordinate, Pressure, Button status, Azimuth, Altitude) are used for feature extraction like other available databases described in Chapter 2 [3], [223]. Figure 3.12, 3.13 and 3.14 show some examples of the trajectories of these captured raw data items for one sample.



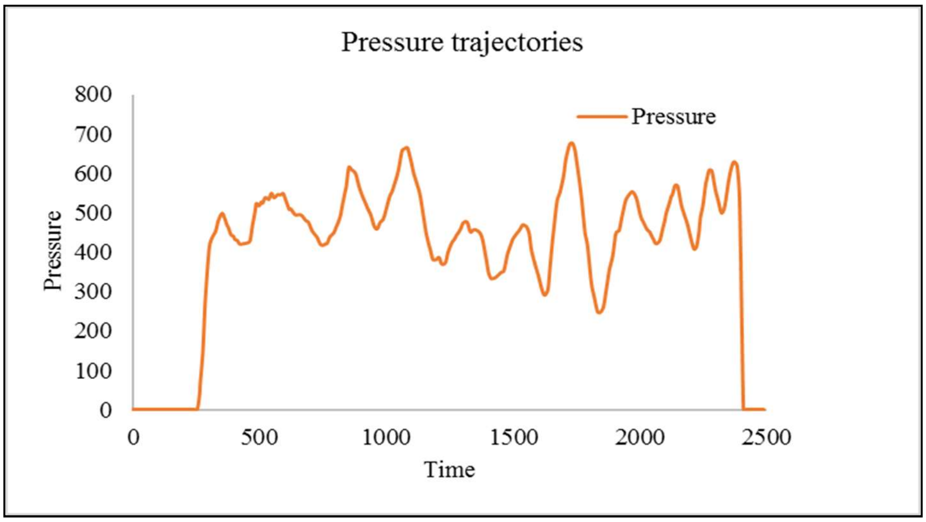


Figure 3.12. Trajectories of x and y pen position (top) and pen pressure (bottom)

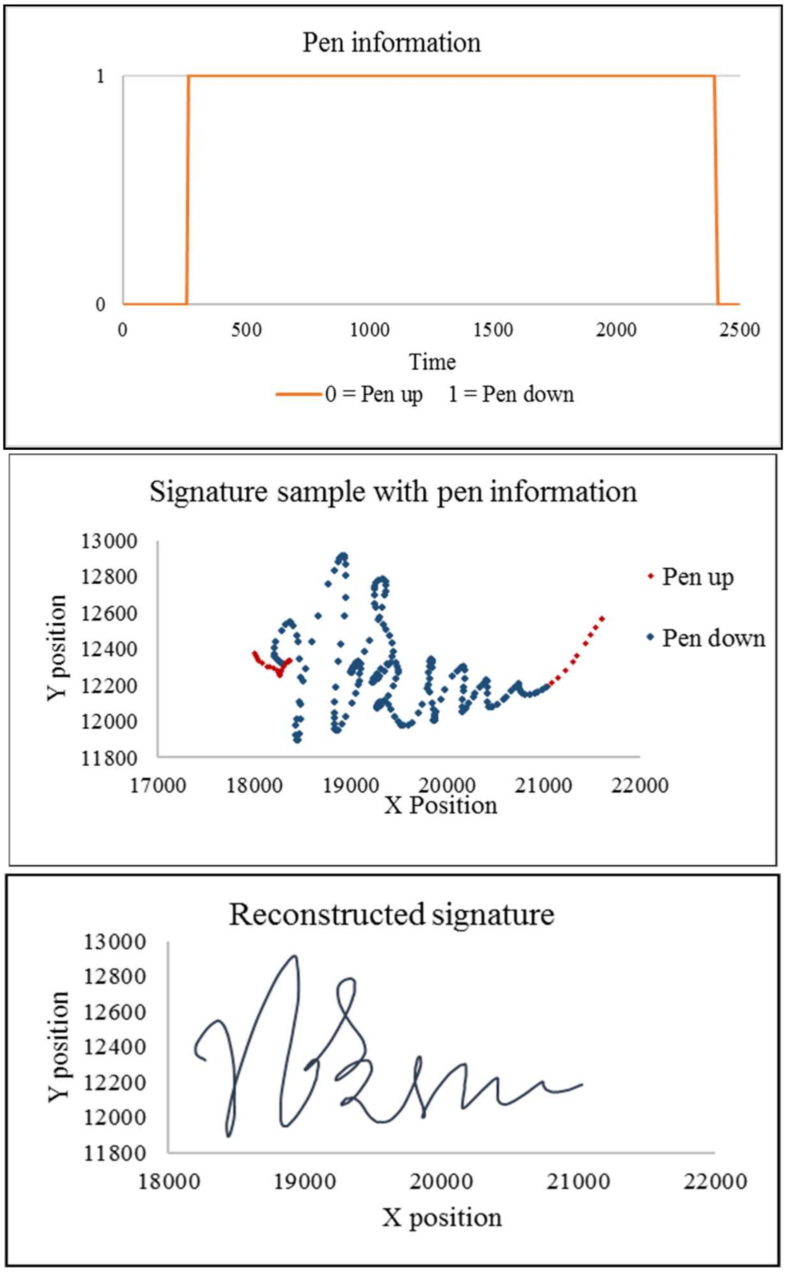


Figure 3.13. Trajectories of pen status (top), pen up and down points (middle) and reconstructed signature from pen down points (bottom)

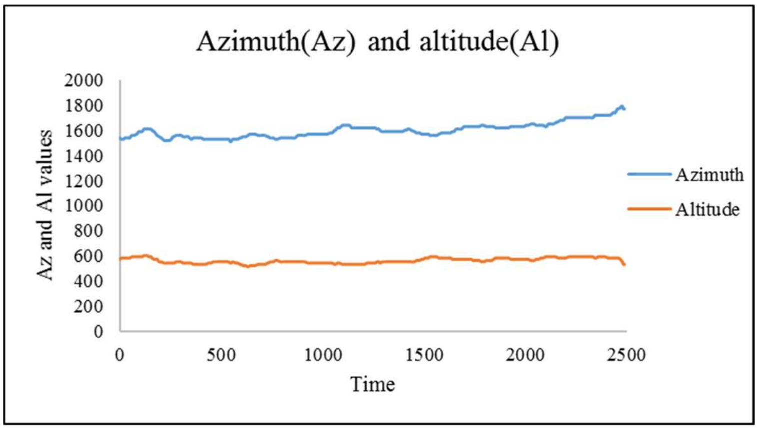


Figure 3.14. Trajectories of pen azimuth and altitude

3.4 Collected data and test subject information

As mentioned earlier, in the recruitment phase the population taking part in this database collection mainly consisted of staff and students of the University of Kent, but also together with some members of the general public.

The following information was obtained from each test subject:

- Name (Family name and First names).
- Gender.
- Date of birth
- Nationality
- First speaking and writing language
- Occupation
- Handedness (Right, Left)

Distribution of the sample population by age group (16- 25, 25 – 40, 40 – 60) is shown in Figure 3.15 and by gender (Male and Female) in Figure 3.16. Figure 3.17 shows the distribution of the sample population according to an individual's tendency to left or right handedness and the distribution according to the subject's ethnic origin is shown in Figure 3.18.

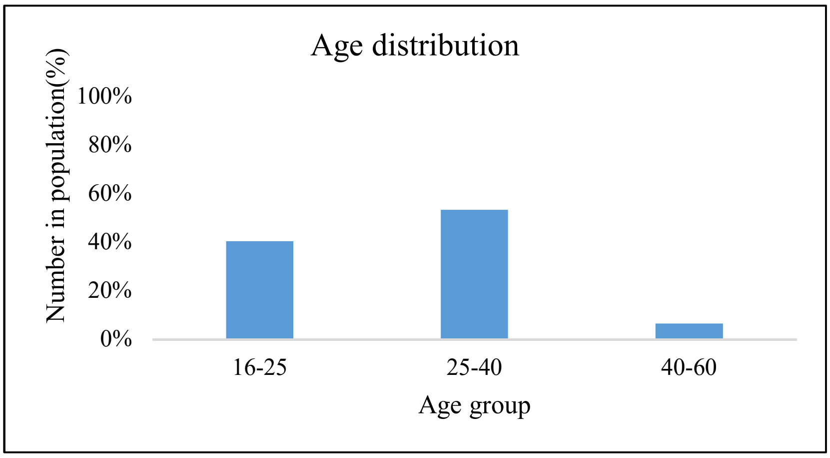


Figure 3.15. Age distribution of the sample population

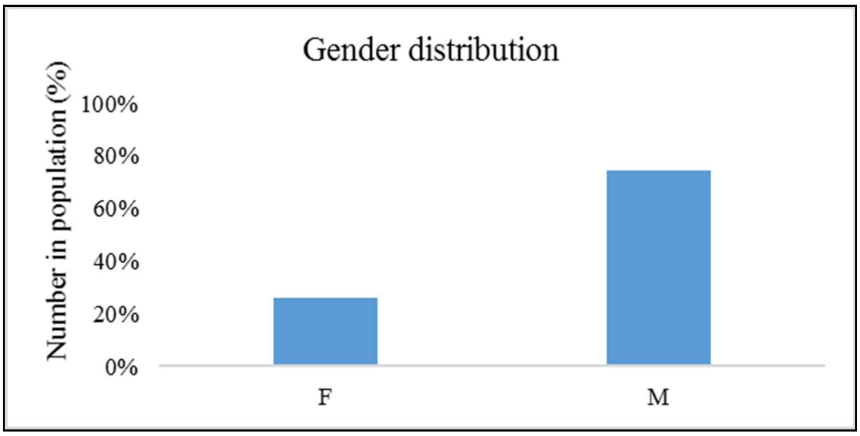


Figure 3.16. Gender distribution of the sample population

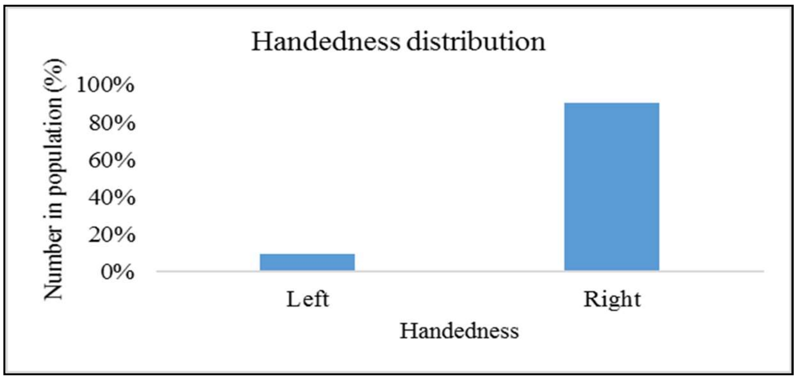


Figure 3.17. Handedness distribution

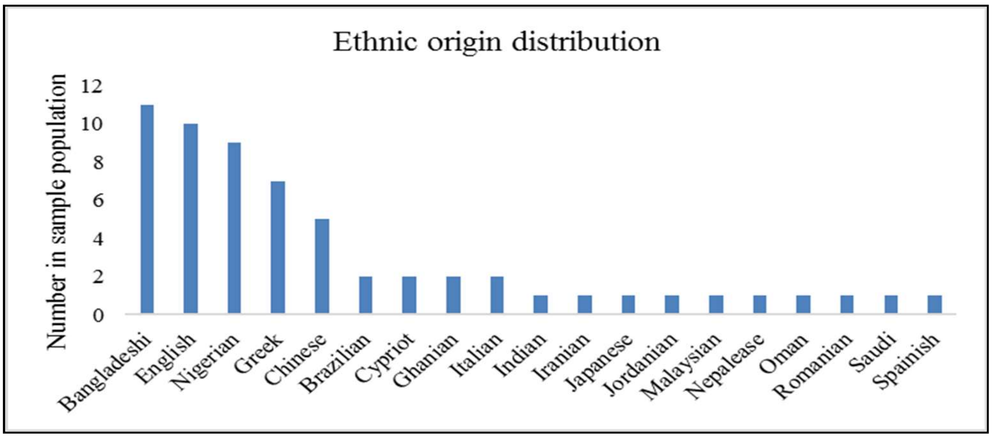


Figure 3.18. Ethnic origin distribution

Data were collected from 62 individual subjects. A total of 4190 signature samples and 1740 handwriting samples were thus donated overall (shown in Figure 3.19). The number of signature samples for each subject varied between 10 and 120. This was dependent on the number of the capture sessions. The number of sessions also varied between 1 and 10. It was originally planned for initially up to 4 collection sessions to be used, with a week's interval between sessions. Later, while the collection process was underway it was increased up to 6 and later to 10 collection sessions to be able to observe the signature development over a longer period of time (reported in Chapter 4 and 6). However, some participants could not complete all the planned sessions and, indeed, some left the process even after the first session. Section 3.5 describes these challenges during data collection and Figure 3.20 shows the distribution of the participants according to the number of sessions in which they took part.

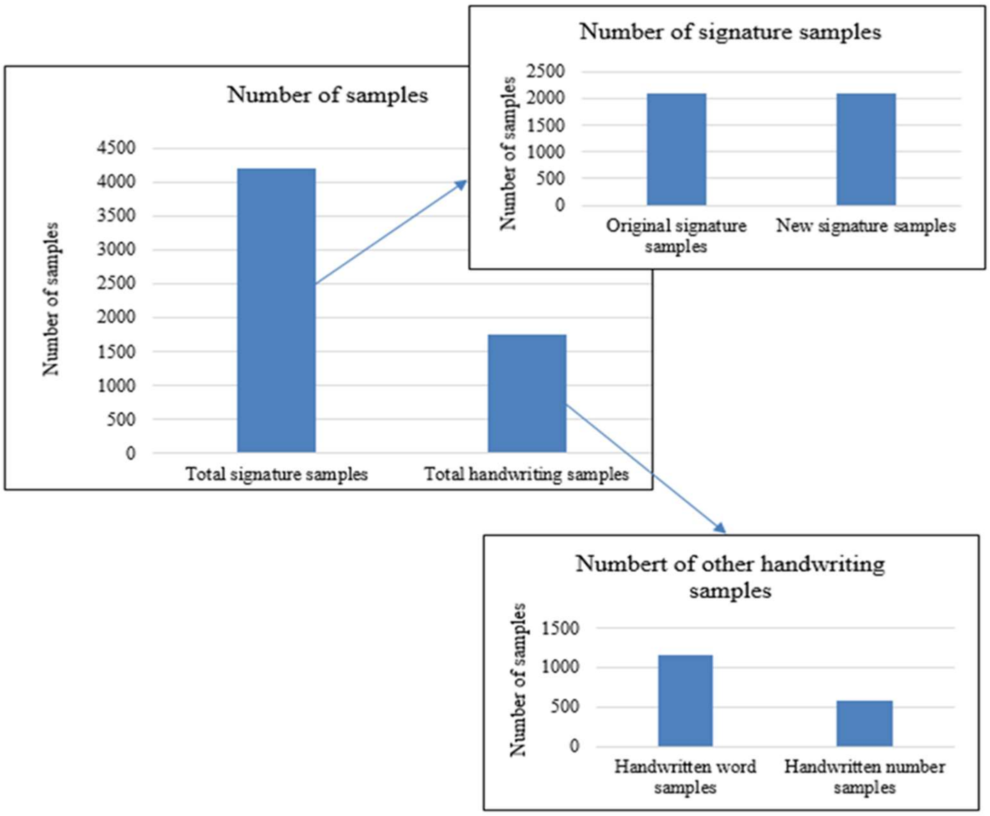


Figure 3.19. Number of collected samples in the database

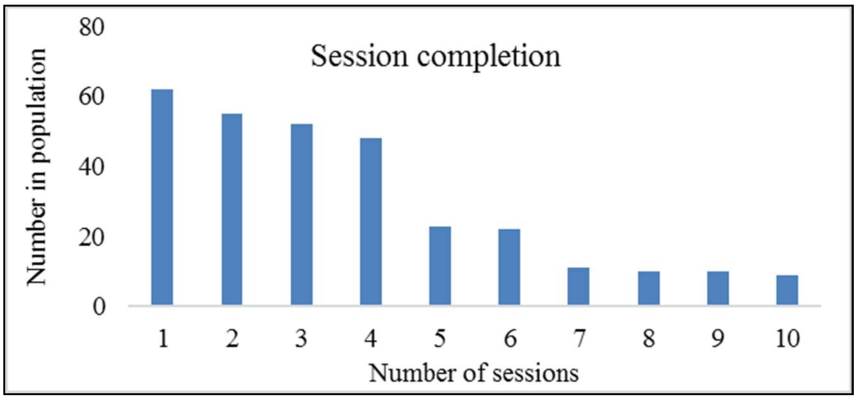


Figure 3.20. Number of sessions completed by the subjects

3.5 Challenges and benefits of the data acquisition

The data acquisition process required a great amount of time and effort because of collecting different handwriting samples for a number of sessions. The nature of this project was completely voluntary and motivational solely by a willingness to support the research effort, without any monetary reward. This posed challenges to recruit people for the data collection. Participants could withdraw at any time during the process between capture sessions, which made it even harder to collect data for all sessions (4 sessions, 6 sessions or 10 sessions depending on the experimental work reported in chapter 4 in this thesis) from the same participant. Figure 3.20 shows an overview of the number of acquisition sessions completed by the participants. A total of 62 participants were involved in this process, where 48 of the participants completed at least 4 sessions and 25 of them left the process after the 4th session, 22 participants continued to the 6th session and 9 of these participants completed up to the 10th session. 7 participants withdrew from the process after just the first session, due to personal reasons.

Since this data acquisition process was such a time-consuming part of the overall work reported in this thesis, it is worth discussing briefly the benefits and, indeed, the necessity of carrying this out. The data acquisition was designed to acquire handwritten signature samples along with other handwriting samples, although the work reported in this thesis mainly focuses on analysis of the handwritten signature. This data collection process involved collecting many samples of a newly developed signature for each subject as well as subject's original signature to study the nature of naturally revoked signature in a situation where the original was compromised. Other available signature databases (described in Chapter 2) consist only of original signatures and some with forgery samples, but not a new signature created by the same signer. This type of database with an original and a new signature created by the same subject was essential for the study and work reported here in this thesis. Also, the number of capture sessions (four, six and up to ten) was important for the experimental work reported in later chapters (Chapter 4, Chapter 5, Chapter 6) in this thesis. The collection of this enhanced signature database was thus important to support a principal strand of the investigation carried out, but was also worthwhile in generating a range of data from each of a number of writers. It therefore provided a good set of signature samples along with other handwriting samples for further analysis. Although the number of participants were not very large due to the challenges discussed above, it can be seen from Figure 3.15 to Figure 3.18 that the participants were from a range of different ethnic background, profession (e.g. students, academics, homemakers, police, businessmen etc.) age, gender which reflect the members of the larger population.

3.6 Conclusion

A brief review of the available data acquisition methods for offline and online capture approaches was initially provided. The data collection protocol and the ethical procedures were described, leading to the construction of a database containing 4190 signature samples and 1740 non-signature handwriting sample, for use in the experiments and applications to be reported in later chapters. The digitizer tablet and software used for this data acquisition was described, while the format of the generated data and the parameters captured by the tablet were also explained. An illustration of the collected data and subject information was provided. Finally, the challenges during data acquisition procedure and the value of this entirely new database collected as part of the project were discussed.

The next chapter will analyse the data collected in this data collection process, and will introduce a detailed study of the concept which we call "natural revocability" in relation to the handwritten signature.

Chapter 4:

Natural Revocability in Handwritten Signature Biometrics

This chapter will present some analysis of the handwritten signature samples collected as described in Chapter 3 to support the initial study of Natural Revocability in handwritten signature biometrics. Section 4.1 will discuss the idea of natural revocability. Section 4.2 will demonstrate the pre-processing of the collected data for natural revocability analysis. Following the data pre-processing some experimental analysis which has been carried out will be reported in section 4.3. Finally, section 4.4 will conclude the chapter.

4.1 Review of revocable biometrics, and the concept of natural revocability

Like most data sources, if not secured, biometric data may be fraudulently obtained or simply stolen, and subsequently misused without a user's consent. Significantly, a compromised biometric is forever compromised if access to the raw information has occurred [13], [14]. Consequently, template protection is a security feature that needs to be addressed in biometric-based authentication systems [8], [49] and such concerns have led researchers to introduce the concept of cancellable biometrics[15], [27], [58], [233]). In this concept a fixed and unchanging biometric template is replaced by a revocable one (i.e. one which can be revoked or changed in the event of compromise), which can be created, for example, through processing by a unidirectional transformation. In the event of compromise, a new biometric template can then be created from the raw data simply by invoking a different transformation. This concept of revocability has been studied extensively with respect to physiological biometrics such as iris, fingerprints etc. [14], [27], [63], [71], [233]–[235]. Some examples of the iris modality include introducing a two-factor scheme using password with iris biometric data for revocable iris template generation. An IrisCode shuffling scheme is employed which increases the Hamming distances between genuine and non-genuine users without changing intra-person distances between genuine users. An Error Correcting Codes (ECC) scheme is also used which corrects more error in the genuine Iris-codes than the non-genuine IrisCodes (this eventually reduces the intra-person Hamming distances between genuine users). Thus, employing the combination of these two schemes increases the verification performance. [59]**.** In a study reported in [235], another approach known as "BioEncoding" is adopted in which a "BioCode", a compact non-invertible bit-string is randomly generated from an original IrisCode and used for user identity verification without affecting the performance when using the original IrisCodes. Gabor filters have also been introduced to achieve the same purpose [236]. In another study reported in [237], a technique based on Steganography where a combination of Huffman Encoding and Discrete Cosine Transformation (DCT) is used to generate cancellable iris template, referred to as “Stego” image or template. Since the transformed Stego image is irreversible, this approach increases the security of the biometric template. In another approach [238], an adaptive Bloom filter-based transform is applied to IrisCodes in order to generate alignment-independent revocable iris template. Over the years a number of revocable biometric schemes have been proposed to increase privacy and security but many at the expense of substantially decreased accuracy in performance. It is shown that [238]–[240] the Bloom filter-based transform not only can protect the template but also maintain the performance. Another scheme named “Indexing-First-One” (IFO) hashing has been introduced by Lai et al. [16] which is inspired by the “Min Hashing” technique that is used for quick estimation of “Jaccard Similarity” between two sets and initially has been used in text matching. To increase the privacy and security the scheme is modified using P-order Hadamard product and modulo threshold function and it has been reported [16] that the IFO hashing scheme maintains recognition accuracy as well as providing privacy and/or security. In the fingerprint modality, reported approaches include introducing various transformation functions for example, polar, Cartesian and surface folding algorithms on the minutiae positions [13], securing the templates by a method called “crypto biometric fuzzy vault framework”. This is implemented by extracting features from the fingerprint, including passwords to provide revocability and then securing the password-hardened revocable templates in a biometric fuzzy vault[241]; extracting the features from the fingerprint minutiae pairs, dividing them into discrete levels, performing histogram equalisation, binarisation and finally generating bit strings [51]. In addition, a process called "Bin-based Quantization (BQ)" is used whereby features such as ridge orientation, frequency, angles between minutiae pairs and total number of minutiae are extracted, eventually revocable templates are generated governed by a secret key providing security[242]. In another approach, a kernel principal components analysis (KPCA) and binarisation techniques are used to generate a fixed-length binary fingerprint template from a variable size and unordered “Multi-line Code” (a minutiae descriptor) fingerprint template.[243]. In the work reported in [157], a non-invertible partial Hadamard transform is applied to generate a complex binary vector representation of the cancellable fingerprint template which makes the retrieval of the original binary vector representation of the template almost impossible. It is also shown that the stochastic distances between binary vectors are preserved after the transformation in this approach.

In this general research area much less attention has been given to behavioural modalities. However, some studies that have been reported in this area of biometrics include using ensemble systems in four ways transformed (Interpolation, BioHashing, BioConvolving and Double Sum) cancellable touchscreen data [244], the process of binding a key to a biometric template for online authentication by encryption at the registration stage and authentication at the final stage through point matching [11], a signer revoking a blind signature to enable looking into the original user activity in cases of foul play through the use of a "magic ink" signature[245]. A biometric hash generation scheme has also been used by connecting the quantised binary strings transformed from the feature vector subsets selected with genetic optimisation [246]. A convolution-based non-invertible transformation approach, BioConvolving, is used in the study reported in [61] to generate revocable signature template, where the signature template is represented as a set of discrete sequences. Another study reports work in the handwritten signature modality using a haptic device (a virtual environment where signers write on a virtual plate)[247] , where a user-specific key is assigned to extracted features and in the event of compromise, a new key is assigned or chosen to cause a different permutation of the features (shuffling of the features).

In the majority of the studies reported above, original signatures are distorted to revoke a new template. However, it is evident that most behavioural modalities, and the handwritten signature in particular, present the possibility of adopting an extremely simple and intuitive strategy for the revocation process. Since such modalities depend entirely on an action carried out by an individual rather than an inherent physiological characteristic, an individual can simply aim to change the execution pattern of whatever action is the source of the biometric data, and which is entirely under that individual's control. We have coined the term "*natural revocability*" for this approach, which has not been studied in relation to biometrics and data revocability hitherto, but which may open up possibilities for revocability strategies proving to be both simple and effective. The simplest example is the consider the handwritten signature. Biometric identification based on the process of signing depends on the assumption (well-supported by observation, intuition and experimentation) that for most people the act of signing becomes sufficiently familiar and ingrained that an individual effortlessly produces multiple examples of his/her signature which are inherently similar. However, such a signature can be revoked simply by not using it anymore, and instead a new signature substituted which, again through repetition, may be assumed over time to become similarly effortlessly reproducible. However, this new approach also raises a number of important and interesting questions, apart from the obvious one about how easy it might be for a familiar to "invent" and personalise a newly formed signature. These include, for example, questions about how long the process of achieving stability in the new signature will take, is stability guaranteed in a signer whose previous signature was stable, what is the likelihood of important features in the original signature carrying over into the new one, and so on. Therefore, this chapter will present some preliminary studies of the phenomenon of natural revocability as a precursor to longer term and more detailed investigations of the potential for natural revocability in behavioural biometrics to be realised as a practical option in the future. We will focus our study on the handwritten signature, recognising that the general principles discussed, although not necessarily all the detail, may also be relevant to other behavioural modalities.

Natural revocability, then, is the term we use to describe the fact that most behavioural biometrics, being under the direct control of the "user", can be created at will in multiple forms. The handwritten signature provides a very good illustrative example [248]. Unlike the case when using a physiological modality such as the fingerprint or iris, the handwritten signature form developed by a particular individual can be discontinued at any point in time, and a new signature invented. This natural revocability potentially offers the opportunity to increase security and privacy while simultaneously avoiding the need for developing alternative and more complex protection techniques. Though the dimensions of handwritten signatures can vary with time [211], [249]–[251] for most people, the fundamental characteristics of the handwritten signature remain relatively constant over a period when written in a given frame [252]. Since it is a voluntary action to change an original signature to a new one, a new biometric can be created easily, thus paralleling closely a user-manipulated password scenario. However, the stability of the form of the signature is generally acquired with repeated use, and it cannot be assumed that all individuals will easily achieve stability with a newly acquired signature. Even if this can be achieved, it is not known whether this is likely to occur on a sufficiently short timescale to make such a change viable in the context of biometric recognition. Thus, in this chapter some fundamental questions will be addressed which need to be considered if the concept of natural revocability is to be exploited as a practical strategy with respect to the signature. Of course, such an investigation also raises a range of other questions of practical importance and a number of these also will be pointed out in this preliminary study.

4.2 Experimental framework for the analysis of natural revocability

This section will describe the experimental set ups and protocols used and some practical details relevant to an exploration of the viability of natural revocability in handwritten signature biometrics.

To study our concept of natural revocability of the handwritten signature it is necessary to establish a database of handwritten samples based first on an individual's current established signature and also of a new signature such as might be adopted should the original need to be withdrawn. So, as described in Chapter 3, under supervised conditions, samples of both were captured from a group of volunteers using a standard pen of familiar style and feel, and an electronic graphics tablet (here a WACOM Intuos-3 tablet with a resolution of 5080 lines per inch) connected to a computer. The system allowed a subject to write normally on a sheet of paper overlaid on the tablet surface, with the pen movement tracked and a representation of the signature stored in the computer in the form of a sequence of time-stamped spatial pen coordinates [146]. Details of the data collection protocols, procedures and storage have been described in Chapter 3.

For this experimental study, two datasets O-RevKent (Original signatures) and N-RevKent (New signatures) of the above-mentioned database – namely Rev-Kent, as designated in Chapter 2, are defined. Following this, a range of commonly used features (listed and specified in Table 2.1, 2.2 and 2.3 in Chapter 2) are extracted from all the signature samples. Extracted features are then normalised and a feature correlation is performed for each dataset. Features with high correlation coefficient are identified for each dataset and discarded from the experiment allowing to use only non-redundant features for the experimental study. Subsequently, the suitability and effectiveness of natural revocability in handwritten signature biometric as a practical option in signature recognition, can be investigated using the defined uncorrelated features by observing how "stability" of the form of the signature changes over a period of time, as this stability in signing (the extent to which the "intrinsic properties of rapid human movements that constitute the basic element of each signature" [198], [253] are reproduced) is a key factor in determining the suitability of the signature for biometric identification.

Furthermore, in order to investigate the characteristics of potential revocability in the signature modality, performances will be analysed by invoking the "biometrics menagerie" notation for individual behaviour which was first introduced by Doddington in the context of speaker recognition [254]. Also, a more practically-oriented test of the viability of the natural revocability concept for the handwritten signature will be performed by evaluating the achievable performance in a recognition/verification scenario.

The next section will report the experimentation in detail with the aim to introduce greater clarity about the potential for *natural revocability* in handwritten signature biometrics to be realised as a practical option, providing some results and analysis.

4.3 Experimentation and results

As described in the previous section, an experimental study is carried out to investigate and explore the potential for *natural revocability* in handwritten signature biometrics to be realised as a practical option. For this experimental study the Rev-Kent database (Data collection is described in Chapter 3) is utilised. Then, the signature samples of O-RevKent and N-RevKent are processed by using the signature processing system as described in Chapter 2.

Hence, recalling the processing chain illustrated in Figure 2.1 (in Chapter 2), 60 features defined in Table 2.1, 2.2 and 2.3(in Chapter 2), are extracted from all samples of the signature database, in the *feature extraction* step. Subsequently, the *feature normalisation* step is carried out and the extracted features are normalised by using the MVN (Mean and Variance Normalisation) technique as defined in

Equation 2.1 (in Chapter 2). Then the Spearman's rank correlation is evaluated between all the extracted and normalised features as explained in Section 2.1.2 (in Chapter 2) in the *feature correlation* step. As a result of this correlation test, 60*60 (3600) correlation values (ρ) are obtained. The correlation results obtained in the experiments are illustrated in Table 4.1 – Table 4.6.

Each value of rho ρ (correlation coefficient) obtained by the evaluation is a number between –1 and 1 that determines the extent to which the feature values are related, with a value closer to zero indicating lower correlation and a value closer to -1 or 1 indicating higher correlation. In this experiment, a value which lies between -0.65 and +0.65 is the criterion used to define "non-correlation". Although the specific experiments are not reported here in detail, this threshold value (0.65) is experimentally determined such that a balance is achieved between performance and the number of features used. Hence, each feature which has at least one value which is not between -0.65 and +0.65 is discarded from the defined feature list. As a result, 25 features are found to be uncorrelated for both original (O-RevKent) and the new (N-RevKent) signature dataset, and the ρ values of 25*25 feature combinations (1 - 5, 8, 10, 13, 14, 17, 19, 22- 24, 27-29, 31, 35, 37, 39, 46, 50, 59, 60) with respect to the O-RevKent and N-RevKent datasets are shown in Table 4.5 - Table 4.6.

Table 4.1. All features (showing 1-30) and their correlation values in the O-RevKent dataset

		Feature correlation values																													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Feature correlation values	1	1	0.5	0.26	-0	0.21	0.21	0.24	0.59	0.7	0.45	0.46	0.63	0.1	0.04	0.13	0.17	0.34	0.33	0.24	0.24	0.33	0.26	-0.2	0.1	0.44	0.33	0.53	0.37	0.26	0.25
	2		1	0.32	-0.3	0.14	0.7	0.74	0.04	0.08	-0.3	-0.3	-0.1	-0	0.05	0.65	0.67	0.08	0.1	-0.2	-0.2	-0	0.19	-0	-0.1	0.81	0.63	0.21	0.08	-0.1	0.11
	3			1	-0.1	0.12	0.05	0.06	-0.1	-0	0.07	0.1	-0	0	-0.1	-0	-0	-0.1	-0.1	0.03	0.01	-0.1	-0.1	0.09	-0	0.18	0.11	0.06	-0	-0	0.09
	4				1	0.07	-0.2	-0.1	0.25	0.01	0.11	0.15	0.3	0.08	0.06	-0.2	-0.1	0.22	0.19	0.07	0.08	0.18	-0.2	-0.1	0.1	-0.1	-0.1	0.39	-0.5	-0.3	0.79
	5					1	-0	-0.1	-0	-0	0.19	0.12	-0	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.2	-0	0.05	0.01	0.17	0	-0	0.14
	6						1	0.95	0.12	0.14	-0.3	-0.3	-0.1	0.03	0.18	0.97	0.93	0.41	0.45	0.09	0.05	0.31	0.14	0.14	-0.2	0.91	0.71	0.05	0.06	-0.1	0.07
	7							1	0.14	0.14	-0.3	-0.4	-0.1	0.02	0.18	0.92	0.97	0.4	0.44	0.04	-0	0.28	0.14	0.14	-0.2	0.91	0.73	0.13	0.03	-0.1	0.15
	8								1	0.88	0.35	0.33	0.91	0.11	0.09	0.06	0.13	0.54	0.56	0.29	0.28	0.51	0.11	-0.1	0.13	0.18	0.15	0.31	0.15	0.19	0.33
	9									1	0.44	0.43	0.84	0.11	0.08	0.07	0.12	0.55	0.58	0.34	0.35	0.54	0.22	0.04	0.13	0.21	0.18	0.28	0.39	0.27	0.11
	10										1	0.94	0.58	0.06	-0	-0.3	-0.3	0.3	0.3	0.57	0.65	0.46	-0	-0	0.08	-0.2	-0.2	0.2	0.25	0.32	0.03
	11											1	0.63	0.09	-0	-0.3	-0.4	0.28	0.27	0.57	0.63	0.43	0.01	-0.1	0.1	-0.2	-0.2	0.22	0.25	0.31	0.06
	12												1	0.14	0.08	-0.1	-0.1	0.52	0.52	0.41	0.41	0.55	0.11	-0.1	0.13	0.02	0.02	0.35	0.19	0.25	0.33
	13													1	0.31	0.03	0.03	0.22	0.1	0.12	0.06	0.15	0.09	-0.1	0.05	0.05	0.03	0.14	0.09	-0	0.1
	14														1	0.19	0.19	0.23	0.21	0.28	0.07	0.08	0.09	-0	-0.1	0.2	0.17	0.07	-0	-0	0.1
	15															1	0.94	0.38	0.43	0.09	0.04	0.29	0.12	0.13	-0.3	0.88	0.69	0.02	0.04	-0.1	0.04
	16																1	0.39	0.44	0.04	-0	0.28	0.11	0.15	-0.3	0.88	0.71	0.1	0.03	-0.1	0.12
	17																	1	0.93	0.53	0.5	0.91	0.04	0.15	-0.1	0.42	0.33	0.28	0.13	0.1	0.27
	18																		1	0.52	0.51	0.88	0.06	0.2	-0.1	0.46	0.36	0.21	0.11	0.1	0.23
	19																			1	0.88	0.61	-0.1	0.08	-0.1	0.14	0.07	0.1	0.23	0.16	-0
	20																				1	0.71	-0.1	0.04	-0.1	0.1	0.02	0.09	0.25	0.13	-0.1
	21																					1	-0	0.1	-0.1	0.33	0.23	0.22	0.18	0.11	0.17
	22																						1	-0.2	0.28	0.16	0.1	-0	0.12	-0.1	-0.1
	23																							1	-0.1	0.07	0.15	-0.2	0.03	0.1	-0.1
	24																								1	-0.2	-0.2	0.09	0.07	-0	0.06
	25																									1	0.77	0.25	0.09	-0.1	0.19
	26																										1	0.24	0.08	-0	0.18
	27																											1	0.17	0.13	0.6
	28																												1	0.29	-0.5
	29																													1	-0.1
	30																														1
	31																														
	32																														
	33																														
	34																														
	35																														
	36																														
	37																														
	38																														
	39																														
	40																														
	41																														
	42																														
	43																														
	44																														
	45																														
	46																														
	47																														
	48																														
	49																														
	50																														
	51																														
	52																														
	53																														
	54																														
	55																														
	56																														
	57																														
	58																														
	59																														
	60																														

Table 4.2. All features (showing 31-60) and their correlation values in the O-RevKent dataset

	Feature correlation values																													
Feature correlation values	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	0.21	0.58	0.73	0.47	0.27	0.2	0.26	0.12	0.12	0.42	0.28	0.12	0.21	0.69	0.41	0.36	0.7	-0.4	-0.7	0.47	-0.2	-0.1	0.24	-0.1	0.17	0.38	0.02	0.21	-0	0.06
2	0.07	0.44	0.51	0.72	-0.3	-0	0.1	0.02	0.08	0.05	-0.1	0.08	0.07	0.5	-0	0.43	0.31	0.03	-0.5	0.61	0.33	0.43	-0.5	0.3	0.61	0.64	0.04	-0.4	-0.1	0.2
3	0.11	0.34	0.41	0.5	-0.3	-0.1	0.03	0.04	-0	-0	-0.2	-0	0.11	0.4	-0.1	0.36	0.22	0.07	-0.4	0.63	0.19	0.26	-0.4	0.18	0.4	0.91	0.55	-0.2	-0	0.17
4	-0	0.48	0.28	-0	0.12	-0	-0.1	0.15	0.03	0.21	0.07	0.03	-0	0.42	0.18	0.29	0.36	-0.2	-0.4	0.07	-0.1	-0.1	0.3	0.02	-0	-0.1	0.11	0.45	-0.1	-0.1
5	-0	-0.1	-0.1	0.08	-0.1	-0.6	-0.5	0.63	0.05	0.05	-0.2	0.05	-0	-0.1	-0	-0	-0.1	0.05	0.1	0.25	0.05	-0	-0.1	0.02	0.06	0.07	0	-0	-0	-0.2
6	-0.1	0.36	0.28	0.77	-0.3	-0	-0.1	-0.1	0.08	-0.1	-0.3	0.08	-0.1	0.35	-0.2	0.4	0.13	0.18	-0.3	0.42	0.84	0.78	-0.6	0.78	0.96	0.16	-0.4	-0.3	-0.1	0.07
7	-0	0.4	0.35	0.78	-0.3	-0	-0.1	-0.1	0.08	-0.1	-0.3	0.08	-0	0.4	-0.2	0.46	0.18	0.18	-0.4	0.39	0.74	0.85	-0.6	0.76	0.95	0.2	-0.4	-0.3	-0.1	0.12
8	0.07	0.43	0.39	0.05	0.35	0.28	0.15	0.01	0.15	0.41	0.26	0.15	0.07	0.45	0.39	0.19	0.54	-0.4	-0.5	-0	-0.4	-0.3	0.62	-0.2	-0.1	-0.1	0.01	0.58	-0	-0
9	0.1	0.33	0.34	-0	0.43	0.34	0.22	-0	0.17	0.38	0.35	0.17	0.1	0.36	0.42	0.11	0.49	-0.4	-0.4	-0.1	-0.5	-0.3	0.7	-0.3	-0.2	-0.2	-0.1	0.57	0.03	-0.1
10	0	0.11	0.03	-0	0.57	0.22	0.07	0.3	0.18	0.56	0.47	0.18	0	0.08	0.59	-0.3	0.37	-0.6	-0.1	0.05	-0.3	-0.4	0.63	-0.2	-0.2	-0.2	0.01	0.7	0.09	-0.1
11	0.01	0.09	0.02	-0.1	0.61	0.3	0.1	0.25	0.18	0.58	0.54	0.18	0.01	0.07	0.63	-0.3	0.38	-0.6	-0.1	0	-0.4	-0.4	0.71	-0.3	-0.3	-0.2	0.03	0.74	0.06	-0.1
12	0.07	0.34	0.28	-0	0.5	0.34	0.17	0.09	0.18	0.53	0.41	0.18	0.07	0.34	0.54	0	0.54	-0.5	-0.3	-0	-0.4	-0.4	0.75	-0.3	-0.2	-0.2	0.01	0.72	-0	-0.1
13	-0.1	-0	-0.1	-0	-0.1	0.04	-0	-0.1	-0	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0	-0.1	0.08	0.06	0.02	0.04	-0	0	0.02	-0	-0	0.03	0.03	-0.1	-0
14	-0	0.13	0.04	0.21	-0.1	0.05	-0	-0.1	0.01	-0.1	-0.1	0.01	-0	0.1	-0.1	0.14	0.02	0.08	-0.1	0.08	0.26	0.25	-0.1	0.25	0.29	-0.1	-0.2	-0	-0	-0.1
15	-0.1	0.36	0.26	0.78	-0.3	-0.1	-0.1	-0.1	0.1	-0.1	-0.3	0.1	-0.1	0.34	-0.2	0.38	0.13	0.16	-0.3	0.39	0.81	0.75	-0.6	0.75	0.93	0.11	-0.4	-0.3	-0.1	0.06
16	-0	0.4	0.33	0.77	-0.3	-0	-0.1	-0.1	0.09	-0.1	-0.3	0.09	-0	0.39	-0.2	0.45	0.17	0.18	-0.4	0.34	0.73	0.83	-0.6	0.75	0.94	0.13	-0.4	-0.3	-0.1	0.09
17	0.03	0.5	0.35	0.3	0.18	0.24	0.04	-0.1	0.15	0.27	0.11	0.15	0.03	0.47	0.24	0.3	0.47	-0.2	-0.5	0.03	0.1	0.13	0.27	0.23	0.3	-0.2	-0.1	0.58	0.07	-0
18	0.03	0.53	0.38	0.36	0.14	0.18	0.04	-0	0.14	0.26	0.08	0.14	0.03	0.5	0.22	0.34	0.47	-0.2	-0.5	0.07	0.15	0.21	0.21	0.27	0.37	-0.1	-0.2	0.52	-0	-0.1
19	0	0.32	0.21	0.3	0.25	0.14	0.09	0.13	0.22	0.3	0.25	0.22	0	0.29	0.3	0.1	0.35	-0.3	-0.3	0.13	-0	-0.1	0.34	0.04	0.16	-0.2	-0.2	0.61	0.07	-0.2
20	-0	0.26	0.18	0.23	0.31	0.05	0.05	0.24	0.18	0.35	0.24	0.18	-0	0.24	0.33	0.04	0.34	-0.3	-0.2	0.12	-0	-0.1	0.39	0.02	0.12	-0.2	-0.2	0.64	0.04	-0.4
21	-0	0.45	0.3	0.27	0.26	0.17	0.05	0.08	0.17	0.36	0.18	0.17	-0	0.42	0.33	0.2	0.47	-0.3	-0.4	0.04	0.06	0.04	0.35	0.17	0.24	-0.2	-0.2	0.65	0.03	-0.2
22	-0.2	-0.1	-0.3	-0.1	-0	0.08	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.2	-0.2	-0.1	-0.1	-0.2	0.08	0.16	0.12	0.13	0.04	0.01	0.08	-0	0.04	0.12	0	-0.1	0.16
23	0.06	0.27	0.31	0.16	-0.2	0.16	0.1	-0.2	-0	-0.1	-0.1	-0	0.06	0.29	-0.1	0.31	0.15	0.08	-0.3	0.19	0.1	0.14	-0.2	0.04	0.11	-0	-0.3	-0.2	-0.1	-0
24	-0.1	0.06	0.06	-0.2	0.18	-0	0.14	0.08	0.07	0.06	0.13	0.07	-0.1	0.06	0.11	-0	0.1	-0.1	-0.1	-0.1	-0.3	-0.3	0.32	-0.3	-0.3	0.03	0.22	0.23	-0	-0
25	0.02	0.51	0.49	0.87	-0.3	-0	-0	-0	0.13	0.03	-0.2	0.13	0.02	0.53	-0.1	0.51	0.31	0.07	-0.5	0.58	0.56	0.59	-0.5	0.55	0.85	0.31	-0.3	-0.2	-0.1	0.07
26	0.6	0.44	0.51	0.7	-0.2	0.02	0.05	-0	0.14	0.06	-0.1	0.14	0.6	0.5	0.01	0.45	0.33	-0	-0.5	0.46	0.32	0.43	-0.3	0.37	0.62	0.3	-0.2	-0.1	-0	0.06
27	0.2	0.86	0.93	0.41	0.09	0.14	0.23	-0	0.05	0.17	0.16	0.05	0.2	0.94	0.21	0.7	0.74	-0.2	-0.9	0.34	-0	0.15	0.06	0.1	0.3	0.32	0.02	0.13	-0.1	0.03
28	0.08	-0.1	0.03	-0.1	0.31	0.26	0.25	-0.1	0.12	-0	0.34	0.12	0.08	-0.1	0.16	-0.1	0.07	-0.2	0.07	-0.3	-0.3	-0.2	0.31	-0.2	-0.2	-0.3	-0.1	0.17	0.3	-0.1
29	0.08	0.07	0.19	-0	0.29	0.17	0.27	0.01	0.04	0.2	0.23	0.04	0.08	0.13	0.23	-0	0.23	-0.2	-0.1	-0.2	-0.2	-0.2	0.29	-0.2	-0.1	-0.1	-0.2	0.12	0	-0
30	0.09	0.78	0.73	0.43	-0.1	-0	0.04	0.12	0.02	0.21	-0	0.02	0.09	0.81	0.12	0.65	0.6	-0.1	-0.8	0.49	0.09	0.15	-0	0.14	0.34	0.33	0.07	0.17	-0.2	0.05
31	1	0.08	0.23	0.06	0.1	0.06	0.09	0.04	0.07	0.09	0.16	0.07	1	0.15	0.14	0.09	0.18	-0.1	-0.1	0.01	-0.2	-0	0.1	-0.1	-0	0.14	0.11	0.05	0.08	0.05
32		1	0.76	0.44	0.12	0.18	0.21	-0	0.07	0.26	0.15	0.07	0.08	0.95	0.26	0.68	0.78	-0.3	-1	0.25	0.07	0.19	0.09	0.15	0.35	0.19	-0.1	0.21	-0.2	0.03
33			1	0.43	0.08	0.11	0.22	0.03	0.07	0.2	0.15	0.07	0.23	0.92	0.22	0.68	0.74	-0.2	-0.9	0.36	-0.1	0.09	0.12	0.04	0.27	0.32	0.01	0.17	-0	0.01
34				1	-0.2	-0	0.01	0.04	0.56	0.07	-0.1	0.56	0.06	0.46	0	0.42	0.3	-0	-0.5	0.49	0.45	0.48	-0.4	0.46	0.73	0.25	-0.3	-0.1	-0	0.03
35					1	0.47	0.44	0.18	0.15	0.67	0.79	0.15	0.1	0.12	0.83	-0.4	0.55	-0.8	-0.1	-0.4	-0.3	-0.3	0.64	-0.2	-0.3	-0.2	-0	0.5	-0	-0
36						1	0.32	-0.6	0.06	0.4	0.41	0.06	0.06	0.17	0.48	-0.2	0.4	-0.5	-0.2	-0.2	-0.1	0.01	0.25	-0	-0	-0.1	-0.1	0.23	0.02	0.2
37							1	0.07	0.02	0.36	0.55	0.02	0.09	0.23	0.51	-0.1	0.47	-0.5	-0.2	-0.2	-0.2	-0.1	0.15	-0.2	-0.1	0.07	0.01	0.04	-0	0.08
38								1	0.1	0.28	0.23	0.1	0.04	-0	0.28	-0.1	0.15	-0.3	0	0.11	-0.2	-0.2	0.14	-0.2	-0.1	0.07	0.05	0.12	-0	-0.2
39									1	0.11	0.14	1	0.07	0.07	0.14	0.02	0.14	-0.1	-0.1	0.01	-0	-0	0.11	0.01	0.08	-0.1	-0.1	0.16	0.07	-0
40										1	0.46	0.11	0.09	0.26	0.89	-0.3	0.68	-0.9	-0.3	-0.1	-0.2	-0.2	0.4	-0.2	-0.1	-0	-0	0.39	-0	-0
41											1	0.14	0.16	0.17	0.77	-0.3	0.54	-0.8	-0.2	-0.3	-0.4	-0.3	0.49	-0.3	-0.3	-0.1	0.04	0.33	0	0
42												1	0.07	0.07	0.14	0.02	0.14	-0.1	-0.1	0.01	-0	-0	0.11	0.01	0.08	-0.1	-0.1	0.16	0.07	-0
43													1	0.15	0.14	0.09	0.18	-0.1	-0.1	0.01	-0.2	-0	0.1	-0.1	-0	0.14	0.11	0.05	0.08	0.05
44														1	0.27	0.72	0.82	-0.3	-1	0.31	0.02	0.16	0.11	0.12	0.34	0.26	-0	0.21	-0.1	0.03
45															1	-0.4	0.74	-1	-0.3	-0.2	-0.3	-0.3	0.48	-0.2	-0.2	-0	-0	0.41	-0	0
46																1	0.27	0.39	-0.7	0.41	0.15	0.29	-0.2	0.22	0.4	0.22	-0.1	-0	-0.1	0.02
47																	1	-0.7	-0.8	0.06	-0.1	-0	0.34	-0	0.14	0.15	-0	0.37	-0.1	0.02
48																		1	0.27	0.2	0.28	0.26	-0.5	0.22	0.17	0.04	0.01	-0.4	0.03	-0
49																			1	-0.3	-0	-0.2	-0.1	-0.1	-0.3	-0.3	0.03	-0.2	0.12	-0
50																				1	0.12	0.1	-0.3	0.06	0.31	0.5	0.13	-0.1	0.02	-0
51																					1	0.82	-0.7	0.88	0.83	-0	-0.4	-0.3	-0.1	0.07
52																						1	-0.6	0.86	0.84	0.07	-0.3	-0.4	-0.1	0.16
53																							1	-0.5	-0.6	-0.4	0.04	0.82	0.04	-0.2
54																								1	0.89	-0	-0.3	-0.2	-0.1	0.04
55																									1	0.11	-0.4	-0.2	-0.1	0.05
56																										1	0.74	-0.3	-0	0.22
57																											1	-0	0.07	0.12
58																												1	0.08	-0.2
59																													1	0.16
60																														1

Table 4.3. All features (showing 1-30) and their correlation values in the N-RevKent dataset

	Feature correlation values																													
Feature correlation values	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	0.02	0.4	0.56	0.4	0.41	0.1	0.05	0.21	0.08	0.4	0.3	0.08	0.02	0.5	0.41	0.14	0.59	-0.4	-0.5	0.3	-0.1	-0	0.45	-0	0.2	0.23	0.01	0.38	0.13	0.05
2	-0	0.3	0.12	0.6	-0.1	-0.1	0.02	0.11	-0	0.02	-0.1	-0	-0	0.23	-0	0.21	0.15	0.02	-0.2	0.44	0.42	0.51	-0.5	0.39	0.61	0.29	-0.2	-0.4	0.1	0.16
3	0.01	0.06	0.17	0.15	-0.2	-0.1	0.01	0.13	0	-0	-0	0	0.01	0.14	-0	0.07	0.08	0	-0.1	0.31	-0.1	-0	-0.1	-0.1	-0	0.94	0.81	-0	0.11	-0
4	0.08	0.58	0.23	-0.1	0.04	-0.1	-0	0.03	0.07	0.01	0.09	0.07	0.08	0.51	0.06	0.39	0.32	-0.1	-0.5	-0.1	-0.3	-0.2	0.31	-0.1	-0.1	-0.1	0.04	0.31	-0.2	-0.1
5	-0	0.05	0.24	0.08	0	-0.7	-0.5	0.69	0.08	0.02	-0.1	0.08	-0	0.13	-0	0.13	0.08	0.02	-0.1	0.22	-0.1	-0.2	0.05	-0.2	-0.1	0.18	0.11	-0	-0	-0.2
6	-0.1	0.32	-0	0.69	-0.2	0.04	-0	-0.1	-0	-0.1	-0.2	-0	-0.1	0.2	-0.1	0.31	0.04	0.15	-0.2	0.32	0.86	0.8	-0.5	0.82	0.95	-0.1	-0.5	-0.2	-0	0.04
7	-0	0.4	0.01	0.69	-0.2	0.06	-0	-0.2	-0	-0.1	-0.2	-0	-0	0.27	-0.2	0.37	0.08	0.15	-0.3	0.29	0.76	0.88	-0.5	0.79	0.93	-0.1	-0.5	-0.3	-0	0.08
8	0.01	0.44	0.27	0.18	0.33	0.2	0.05	-0	0.06	0.31	0.16	0.06	0.01	0.41	0.29	0.17	0.44	-0.3	-0.4	-0	-0	0.03	0.54	0.15	0.2	-0.1	-0.1	0.54	-0.1	-0
9	-0	0.29	0.35	0.21	0.35	0.25	0	-0	0.07	0.33	0.17	0.07	-0	0.35	0.31	0.12	0.43	-0.3	-0.4	0.08	-0	0.01	0.59	0.12	0.2	-0.1	-0.1	0.59	0.04	-0
10	0	-0	0.39	-0	0.58	0.11	0.06	0.31	0.17	0.53	0.43	0.17	0	0.17	0.55	-0.2	0.45	-0.6	-0.2	-0	-0.3	-0.4	0.73	-0.2	-0.2	0.03	0.16	0.74	0.01	-0.1
11	0.03	0.01	0.41	-0.1	0.63	0.21	0.13	0.26	0.17	0.56	0.54	0.17	0.03	0.2	0.63	-0.3	0.51	-0.6	-0.2	-0	-0.4	-0.5	0.77	-0.3	-0.3	0.07	0.21	0.74	0.05	-0.1
12	0.02	0.37	0.37	0.07	0.52	0.28	0.12	0.03	0.09	0.46	0.36	0.09	0.02	0.42	0.48	0.03	0.56	-0.5	-0.4	-0.1	-0.2	-0.2	0.73	-0	0.01	-0.1	-0	0.71	-0	-0
13	-0	0.09	0.13	0.07	0.03	0.05	0.01	-0	0.05	0.02	0.05	0.05	-0	0.12	0.02	0.1	0.08	-0	-0.1	0.05	-0	-0	0.12	-0	0.02	-0	-0	0.15	0.07	0.01
14	0.01	0.12	0.04	0.18	-0	0.06	0.05	-0.1	0.06	-0	0.03	0.06	0.01	0.09	-0	0.13	0.03	0.03	-0.1	0.12	0.11	0.11	0.01	0.11	0.16	-0.1	-0.2	0.08	-0	-0
15	-0.1	0.29	-0.1	0.68	-0.3	0.01	-0	-0.1	-0	-0.1	-0.2	-0	-0.1	0.16	-0.2	0.3	0	0.17	-0.2	0.31	0.84	0.78	-0.5	0.78	0.91	-0.2	-0.5	-0.3	-0	0.02
16	-0	0.37	-0	0.68	-0.2	0.06	-0	-0.2	-0	-0.1	-0.2	-0	-0	0.24	-0.2	0.35	0.06	0.16	-0.2	0.23	0.76	0.87	-0.5	0.79	0.92	-0.2	-0.5	-0.3	-0	0.05
17	-0	0.46	0.29	0.38	0.18	0.25	0.05	-0.1	0.08	0.24	0.12	0.08	-0	0.46	0.2	0.28	0.41	-0.2	-0.5	0.08	0.27	0.26	0.28	0.38	0.45	-0.3	-0.3	0.53	0.02	-0
18	-0	0.45	0.22	0.4	0.16	0.22	0.03	-0.1	0.07	0.25	0.08	0.07	-0	0.42	0.2	0.26	0.38	-0.2	-0.4	0.09	0.31	0.3	0.25	0.41	0.49	-0.3	-0.3	0.5	-0.1	-0.1
19	-0	0.07	0.23	0.21	0.29	0.18	0.09	0.04	0.18	0.32	0.25	0.18	-0	0.15	0.33	-0	0.3	-0.3	-0.1	0.06	-0	-0.1	0.42	0.07	0.13	-0.1	-0.1	0.63	-0	-0.2
20	-0	0.04	0.23	0.17	0.34	0.15	0.09	0.09	0.17	0.38	0.26	0.17	-0	0.12	0.38	-0.1	0.32	-0.4	-0.1	0.06	-0	-0.2	0.45	0.04	0.08	-0.1	-0.1	0.65	-0	-0.3
21	-0	0.35	0.26	0.32	0.27	0.24	0.08	-0.1	0.1	0.34	0.18	0.1	-0	0.37	0.31	0.14	0.41	-0.3	-0.4	0.05	0.19	0.14	0.36	0.31	0.36	-0.3	-0.3	0.61	-0	-0.1
22	-0.1	0.02	-0.1	0.05	-0.1	0.08	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0	-0.1	0.08	-0.1	0.11	0.02	0.25	0.1	0.11	0.02	0.07	0.11	-0	-0.1	-0.1	0.06	0.14
23	0.13	0.01	-0.1	0.03	-0.2	0.2	0.06	-0.2	-0	-0.1	-0	-0	0.13	0	-0.1	0.07	-0.1	0.07	-0	-0	0.16	0.2	-0.1	0.18	0.14	-0	-0	0.06	-0.2	-0
24	0.05	0.06	0.03	-0.1	0.1	0.04	0.06	0.02	0.07	0.05	0.12	0.07	0.05	0.04	0.08	-0	0.09	-0.1	-0	0.01	-0.3	-0.2	0.2	-0.2	-0.2	0.03	0.08	0.06	0.03	0.1
25	-0	0.43	0.16	0.8	-0.2	-0	-0	-0	0.04	0.01	-0.1	0.04	-0	0.35	-0.1	0.39	0.18	0.07	-0.3	0.5	0.6	0.64	-0.4	0.61	0.84	-0	-0.4	-0.1	0.06	0.03
26	0.51	0.38	0.16	0.61	-0.1	0.01	0.01	-0	0.05	0.03	-0.1	0.05	0.51	0.32	-0	0.34	0.19	0.03	-0.3	0.33	0.49	0.54	-0.3	0.51	0.67	-0	-0.3	-0.1	0.03	0.03
27	0.18	0.7	0.84	0.25	0.24	-0.1	0.06	0.23	0.09	0.18	0.24	0.09	0.18	0.86	0.24	0.52	0.67	-0.2	-0.9	0.13	-0.2	-0.1	0.35	-0.1	0.08	0.04	-0.1	0.28	0.13	0.01
28	0	-0.2	0.31	0.12	0.3	0.2	0.07	0.02	0.07	0.18	0.23	0.07	0	0.02	0.23	-0.1	0.18	-0.2	-0	0.02	0.01	-0	0.27	0.04	0.08	-0.1	-0.1	0.25	0.32	-0
29	0.13	-0	0.31	-0.1	0.42	0.18	0.23	0.1	-0	0.33	0.29	-0	0.13	0.13	0.36	-0.1	0.32	-0.4	-0.1	-0.3	-0.1	-0.1	0.32	-0	-0	0.02	0.02	0.27	0.02	0
30	0.15	0.78	0.43	0.17	0.03	-0.1	-0	0.13	0.05	0.09	0.07	0.05	0.15	0.74	0.09	0.54	0.48	-0.1	-0.7	0.1	-0.1	0.02	0.16	0.02	0.1	0.06	-0	0.14	-0.1	0.05
31	1	0.11	0.19	0.01	0.09	0.04	0.11	0.02	0.11	0.1	0.13	0.11	1	0.17	0.13	0.06	0.18	-0.1	-0.2	-0.1	-0.1	-0	0.06	-0	-0	0.03	0.02	0.03	0.04	0.02
32		1	0.44	0.37	0.07	0.02	0.07	0.04	0.05	0.15	0.07	0.05	0.11	0.89	0.15	0.61	0.62	-0.1	-0.9	0.11	0.07	0.22	0.12	0.23	0.36	-0	-0.2	0.14	-0.1	0.03
33			1	0.2	0.3	-0	-0	0.26	0.12	0.26	0.29	0.12	0.19	0.77	0.31	0.4	0.67	-0.3	-0.8	0.15	-0.3	-0.2	0.46	-0.1	0	0.15	0.07	0.41	0.19	-0
34				1	-0.1	-0	-0.1	0.03	0.6	0.05	-0	0.6	0.01	0.32	0.01	0.31	0.22	-0	-0.3	0.42	0.4	0.44	-0.2	0.44	0.65	-0	-0.3	-0	0.07	0.01
35					1	0.35	0.45	0.26	0.1	0.73	0.76	0.1	0.09	0.18	0.88	-0.5	0.68	-0.9	-0.2	-0.5	-0.3	-0.3	0.57	-0.1	-0.2	-0.1	-0	0.47	-0	0.03
36						1	0.42	-0.7	-0	0.33	0.35	-0	0.04	0.02	0.38	-0.3	0.25	-0.4	-0	-0.3	0.08	0.11	0.2	0.15	0.1	-0.2	-0.1	0.24	0.07	0.22
37							1	-0	-0	0.41	0.52	-0	0.11	0.05	0.54	-0.4	0.39	-0.5	-0.1	-0.3	-0.1	-0	0.05	-0	-0	0	-0	0.03	-0.1	0.06
38								1	0.09	0.25	0.2	0.09	0.02	0.12	0.28	-0.1	0.27	-0.3	-0.1	0.09	-0.2	-0.3	0.12	-0.3	-0.2	0.19	0.15	0.01	-0.1	-0.2
39									1	0.08	0.14	1	0.11	0.08	0.12	0.01	0.13	-0.1	-0.1	0.06	-0.1	-0.1	0.13	-0.1	-0	-0	-0	0.15	0.02	-0
40										1	0.42	0.08	0.1	0.25	0.88	-0.4	0.71	-0.9	-0.2	-0.3	-0.1	-0.2	0.39	-0.1	-0	0.01	-0	0.37	0.01	0.01
41											1	0.14	0.13	0.18	0.76	-0.4	0.6	-0.8	-0.2	-0.3	-0.3	-0.3	0.43	-0.2	-0.2	-0	0.04	0.33	0.01	0.06
42												1	0.11	0.08	0.12	0.01	0.13	-0.1	-0.1	0.06	-0.1	-0.1	0.13	-0.1	-0	-0	-0	0.15	0.02	-0
43													1	0.17	0.13	0.06	0.18	-0.1	-0.2	-0.1	-0.1	-0	0.06	-0	-0	0.03	0.02	0.03	0.04	0.02
44														1	0.26	0.6	0.75	-0.3	-1	0.11	-0.1	0.08	0.28	0.11	0.24	0.07	-0.1	0.29	0.03	0.01
45															1	-0.5	0.8	-1	-0.3	-0.3	-0.2	-0.3	0.46	-0.1	-0.1	0.01	0	0.39	-0	0.02
46																1	0.01	0.53	-0.6	0.41	0.15	0.27	-0.1	0.22	0.31	-0	-0.1	0.01	0.02	-0
47																	1	-0.8	-0.8	-0.1	-0.2	-0.1	0.45	-0	0.09	0.06	-0.1	0.4	0.02	-0
48																		1	0.26	0.33	0.24	0.26	-0.5	0.13	0.1	-0	-0	-0.4	0	-0
49																			1	-0.1	0.05	-0.1	-0.3	-0.1	-0.2	-0.1	0.07	-0.3	-0	-0
50																				1	0.07	0.02	-0.1	-0	0.19	0.2	0.03	-0.1	0.24	-0
51																					1	0.8	-0.6	0.88	0.84	-0.2	-0.4	-0.3	-0.1	0.05
52																						1	-0.6	0.84	0.83	-0.2	-0.4	-0.4	-0.1	0.1
53																							1	-0.4	-0.4	-0.1	0.14	0.87	0.03	-0.1
54																								1	0.93	-0.3	-0.5	-0.1	-0.1	0.06
55																									1	-0.2	-0.5	-0.1	-0	0.04
56																										1	0.88	-0.2	0.12	0.01
57																											1	0.07	0.07	-0.1
58																												1	0.04	-0.1
59																													1	0.25
60																														1

Table 4.4. All features (showing 31-60) and their correlation values in the N-RevKent

	Feature correlation values																													
Feature correlation values	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	0.02	0.4	0.56	0.4	0.41	0.1	0.05	0.21	0.08	0.4	0.3	0.08	0.02	0.5	0.41	0.14	0.59	-0.4	-0.5	0.3	-0.1	-0	0.45	-0	0.2	0.23	0.01	0.38	0.13	0.05
2	-0	0.3	0.12	0.6	-0.1	-0.1	0.02	0.11	-0	0.02	-0.1	-0	-0	0.23	-0	0.21	0.15	0.02	-0.2	0.44	0.42	0.51	-0.5	0.39	0.61	0.29	-0.2	-0.4	0.1	0.16
3	0.01	0.06	0.17	0.15	-0.2	-0.1	0.01	0.13	0	-0	-0	0	0.01	0.14	-0	0.07	0.08	0	-0.1	0.31	-0.1	-0	-0.1	-0.1	-0	0.94	0.81	-0	0.11	-0
4	0.08	0.58	0.23	-0.1	0.04	-0.1	-0	0.03	0.07	0.01	0.09	0.07	0.08	0.51	0.06	0.39	0.32	-0.1	-0.5	-0.1	-0.3	-0.2	0.31	-0.1	-0.1	-0.1	0.04	0.31	-0.2	-0.1
5	-0	0.05	0.24	0.08	0	-0.7	-0.5	0.69	0.08	0.02	-0.1	0.08	-0	0.13	-0	0.13	0.08	0.02	-0.1	0.22	-0.1	-0.2	0.05	-0.2	-0.1	0.18	0.11	-0	-0	-0.2
6	-0.1	0.32	-0	0.69	-0.2	0.04	-0	-0.1	-0	-0.1	-0.2	-0	-0.1	0.2	-0.1	0.31	0.04	0.15	-0.2	0.32	0.86	0.8	-0.5	0.82	0.95	-0.1	-0.5	-0.2	-0	0.04
7	-0	0.4	0.01	0.69	-0.2	0.06	-0	-0.2	-0	-0.1	-0.2	-0	-0	0.27	-0.2	0.37	0.08	0.15	-0.3	0.29	0.76	0.88	-0.5	0.79	0.93	-0.1	-0.5	-0.3	-0	0.08
8	0.01	0.44	0.27	0.18	0.33	0.2	0.05	-0	0.06	0.31	0.16	0.06	0.01	0.41	0.29	0.17	0.44	-0.3	-0.4	-0	-0	0.03	0.54	0.15	0.2	-0.1	-0.1	0.54	-0.1	-0
9	-0	0.29	0.35	0.21	0.35	0.25	0	-0	0.07	0.33	0.17	0.07	-0	0.35	0.31	0.12	0.43	-0.3	-0.4	0.08	-0	0.01	0.59	0.12	0.2	-0.1	-0.1	0.59	0.04	-0
10	0	-0	0.39	-0	0.58	0.11	0.06	0.31	0.17	0.53	0.43	0.17	0	0.17	0.55	-0.2	0.45	-0.6	-0.2	-0	-0.3	-0.4	0.73	-0.2	-0.2	0.03	0.16	0.74	0.01	-0.1
11	0.03	0.01	0.41	-0.1	0.63	0.21	0.13	0.26	0.17	0.56	0.54	0.17	0.03	0.2	0.63	-0.3	0.51	-0.6	-0.2	-0	-0.4	-0.5	0.77	-0.3	-0.3	0.07	0.21	0.74	0.05	-0.1
12	0.02	0.37	0.37	0.07	0.52	0.28	0.12	0.03	0.09	0.46	0.36	0.09	0.02	0.42	0.48	0.03	0.56	-0.5	-0.4	-0.1	-0.2	-0.2	0.73	-0	0.01	-0.1	-0	0.71	-0	-0
13	-0	0.09	0.13	0.07	0.03	0.05	0.01	-0	0.05	0.02	0.05	0.05	-0	0.12	0.02	0.1	0.08	-0	-0.1	0.05	-0	-0	0.12	-0	0.02	-0	-0	0.15	0.07	0.01
14	0.01	0.12	0.04	0.18	-0	0.06	0.05	-0.1	0.06	-0	0.03	0.06	0.01	0.09	-0	0.13	0.03	0.03	-0.1	0.12	0.11	0.11	0.01	0.11	0.16	-0.1	-0.2	0.08	-0	-0
15	-0.1	0.29	-0.1	0.68	-0.3	0.01	-0	-0.1	-0	-0.1	-0.2	-0	-0.1	0.16	-0.2	0.3	0	0.17	-0.2	0.31	0.84	0.78	-0.5	0.78	0.91	-0.2	-0.5	-0.3	-0	0.02
16	-0	0.37	-0	0.68	-0.2	0.06	-0	-0.2	-0	-0.1	-0.2	-0	-0	0.24	-0.2	0.35	0.06	0.16	-0.2	0.23	0.76	0.87	-0.5	0.79	0.92	-0.2	-0.5	-0.3	-0	0.05
17	-0	0.46	0.29	0.38	0.18	0.25	0.05	-0.1	0.08	0.24	0.12	0.08	-0	0.46	0.2	0.28	0.41	-0.2	-0.5	0.08	0.27	0.26	0.28	0.38	0.45	-0.3	-0.3	0.53	0.02	-0
18	-0	0.45	0.22	0.4	0.16	0.22	0.03	-0.1	0.07	0.25	0.08	0.07	-0	0.42	0.2	0.26	0.38	-0.2	-0.4	0.09	0.31	0.3	0.25	0.41	0.49	-0.3	-0.3	0.5	-0.1	-0.1
19	-0	0.07	0.23	0.21	0.29	0.18	0.09	0.04	0.18	0.32	0.25	0.18	-0	0.15	0.33	-0	0.3	-0.3	-0.1	0.06	-0	-0.1	0.42	0.07	0.13	-0.1	-0.1	0.63	-0	-0.2
20	-0	0.04	0.23	0.17	0.34	0.15	0.09	0.09	0.17	0.38	0.26	0.17	-0	0.12	0.38	-0.1	0.32	-0.4	-0.1	0.06	-0	-0.2	0.45	0.04	0.08	-0.1	-0.1	0.65	-0	-0.3
21	-0	0.35	0.26	0.32	0.27	0.24	0.08	-0.1	0.1	0.34	0.18	0.1	-0	0.37	0.31	0.14	0.41	-0.3	-0.4	0.05	0.19	0.14	0.36	0.31	0.36	-0.3	-0.3	0.61	-0	-0.1
22	-0.1	0.02	-0.1	0.05	-0.1	0.08	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0	-0.1	0.08	-0.1	0.11	0.02	0.25	0.1	0.11	0.02	0.07	0.11	-0	-0.1	-0.1	0.06	0.14
23	0.13	0.01	-0.1	0.03	-0.2	0.2	0.06	-0.2	-0	-0.1	-0	-0	0.13	0	-0.1	0.07	-0.1	0.07	-0	-0	0.16	0.2	-0.1	0.18	0.14	-0	-0	0.06	-0.2	-0
24	0.05	0.06	0.03	-0.1	0.1	0.04	0.06	0.02	0.07	0.05	0.12	0.07	0.05	0.04	0.08	-0	0.09	-0.1	-0	0.01	-0.3	-0.2	0.2	-0.2	-0.2	0.03	0.08	0.06	0.03	0.1
25	-0	0.43	0.16	0.8	-0.2	-0	-0	-0	0.04	0.01	-0.1	0.04	-0	0.35	-0.1	0.39	0.18	0.07	-0.3	0.5	0.6	0.64	-0.4	0.61	0.84	-0	-0.4	-0.1	0.06	0.03
26	0.51	0.38	0.16	0.61	-0.1	0.01	0.01	-0	0.05	0.03	-0.1	0.05	0.51	0.32	-0	0.34	0.19	0.03	-0.3	0.33	0.49	0.54	-0.3	0.51	0.67	-0	-0.3	-0.1	0.03	0.03
27	0.18	0.7	0.84	0.25	0.24	-0.1	0.06	0.23	0.09	0.18	0.24	0.09	0.18	0.86	0.24	0.52	0.67	-0.2	-0.9	0.13	-0.2	-0.1	0.35	-0.1	0.08	0.04	-0.1	0.28	0.13	0.01
28	0	-0.2	0.31	0.12	0.3	0.2	0.07	0.02	0.07	0.18	0.23	0.07	0	0.02	0.23	-0.1	0.18	-0.2	-0	0.02	0.01	-0	0.27	0.04	0.08	-0.1	-0.1	0.25	0.32	-0
29	0.13	-0	0.31	-0.1	0.42	0.18	0.23	0.1	-0	0.33	0.29	-0	0.13	0.13	0.36	-0.1	0.32	-0.4	-0.1	-0.3	-0.1	-0.1	0.32	-0	-0	0.02	0.02	0.27	0.02	0
30	0.15	0.78	0.43	0.17	0.03	-0.1	-0	0.13	0.05	0.09	0.07	0.05	0.15	0.74	0.09	0.54	0.48	-0.1	-0.7	0.1	-0.1	0.02	0.16	0.02	0.1	0.06	-0	0.14	-0.1	0.05
31	1	0.11	0.19	0.01	0.09	0.04	0.11	0.02	0.11	0.1	0.13	0.11	1	0.17	0.13	0.06	0.18	-0.1	-0.2	-0.1	-0.1	-0	0.06	-0	-0	0.03	0.02	0.03	0.04	0.02
32		1	0.44	0.37	0.07	0.02	0.07	0.04	0.05	0.15	0.07	0.05	0.11	0.89	0.15	0.61	0.62	-0.1	-0.9	0.11	0.07	0.22	0.12	0.23	0.36	-0	-0.2	0.14	-0.1	0.03
33			1	0.2	0.3	-0	-0	0.26	0.12	0.26	0.29	0.12	0.19	0.77	0.31	0.4	0.67	-0.3	-0.8	0.15	-0.3	-0.2	0.46	-0.1	0	0.15	0.07	0.41	0.19	-0
34				1	-0.1	-0	-0.1	0.03	0.6	0.05	-0	0.6	0.01	0.32	0.01	0.31	0.22	-0	-0.3	0.42	0.4	0.44	-0.2	0.44	0.65	-0	-0.3	-0	0.07	0.01
35					1	0.35	0.45	0.26	0.1	0.73	0.76	0.1	0.09	0.18	0.88	-0.5	0.68	-0.9	-0.2	-0.5	-0.3	-0.3	0.57	-0.1	-0.2	-0.1	-0	0.47	-0	0.03
36						1	0.42	-0.7	-0	0.33	0.35	-0	0.04	0.02	0.38	-0.3	0.25	-0.4	-0	-0.3	0.08	0.11	0.2	0.15	0.1	-0.2	-0.1	0.24	0.07	0.22
37							1	-0	-0	0.41	0.52	-0	0.11	0.05	0.54	-0.4	0.39	-0.5	-0.1	-0.3	-0.1	-0	0.05	-0	-0	0	-0	0.03	-0.1	0.06
38								1	0.09	0.25	0.2	0.09	0.02	0.12	0.28	-0.1	0.27	-0.3	-0.1	0.09	-0.2	-0.3	0.12	-0.3	-0.2	0.19	0.15	0.01	-0.1	-0.2
39									1	0.08	0.14	1	0.11	0.08	0.12	0.01	0.13	-0.1	-0.1	0.06	-0.1	-0.1	0.13	-0.1	-0	-0	-0	0.15	0.02	-0
40										1	0.42	0.08	0.1	0.25	0.88	-0.4	0.71	-0.9	-0.2	-0.3	-0.1	-0.2	0.39	-0.1	-0	0.01	-0	0.37	0.01	0.01
41											1	0.14	0.13	0.18	0.76	-0.4	0.6	-0.8	-0.2	-0.3	-0.3	-0.3	0.43	-0.2	-0.2	-0	0.04	0.33	0.01	0.06
42												1	0.11	0.08	0.12	0.01	0.13	-0.1	-0.1	0.06	-0.1	-0.1	0.13	-0.1	-0	-0	-0	0.15	0.02	-0
43													1	0.17	0.13	0.06	0.18	-0.1	-0.2	-0.1	-0.1	-0	0.06	-0	-0	0.03	0.02	0.03	0.04	0.02
44														1	0.26	0.6	0.75	-0.3	-1	0.11	-0.1	0.08	0.28	0.11	0.24	0.07	-0.1	0.29	0.03	0.01
45															1	-0.5	0.8	-1	-0.3	-0.3	-0.2	-0.3	0.46	-0.1	-0.1	0.01	0	0.39	-0	0.02
46																1	0.01	0.53	-0.6	0.41	0.15	0.27	-0.1	0.22	0.31	-0	-0.1	0.01	0.02	-0
47																	1	-0.8	-0.8	-0.1	-0.2	-0.1	0.45	-0	0.09	0.06	-0.1	0.4	0.02	-0
48																		1	0.26	0.33	0.24	0.26	-0.5	0.13	0.1	-0	-0	-0.4	0	-0
49																			1	-0.1	0.05	-0.1	-0.3	-0.1	-0.2	-0.1	0.07	-0.3	-0	-0
50																				1	0.07	0.02	-0.1	-0	0.19	0.2	0.03	-0.1	0.24	-0
51																					1	0.8	-0.6	0.88	0.84	-0.2	-0.4	-0.3	-0.1	0.05
52																						1	-0.6	0.84	0.83	-0.2	-0.4	-0.4	-0.1	0.1
53																							1	-0.4	-0.4	-0.1	0.14	0.87	0.03	-0.1
54																								1	0.93	-0.3	-0.5	-0.1	-0.1	0.06
55																									1	-0.2	-0.5	-0.1	-0	0.04
56																										1	0.88	-0.2	0.12	0.01
57																											1	0.07	0.07	-0.1
58																												1	0.04	-0.1
59																													1	0.25
60																														1

Table 4.5. Uncorrelated features and their correlation values in the O-RevKent dataset

Feature correlation values	Feature correlation values																								
	1	2	4	5	8	10	13	14	17	19	22	23	24	27	28	29	31	35	36	37	38	39	50	59	60
1	1	0.6	0.02	-0	0.46	0.29	-0.1	-0	0.27	0.26	-0.1	0.15	-0	0.65	0.09	0.26	0.21	0.27	0.2	0.26	0.12	0.12	0.47	-0	0.06
2		1	-0.1	0.07	-0	-0.2	-0	0.12	0.06	0.01	-0	0.21	-0.1	0.5	-0.2	-0	0.07	-0.3	-0	0.1	0.02	0.08	0.61	-0.1	0.2
4			1	0.17	0.33	0.32	-0.1	0.03	0.4	0.3	0.15	-0	0.22	0.39	-0.5	-0.3	-0	0.12	-0	-0.1	0.15	0.03	0.07	-0.1	-0.1
5				1	-0.1	0.14	-0.1	-0.1	-0.1	0	-0	-0.2	-0	-0.1	-0.3	-0.2	-0	-0.1	-0.6	-0.5	0.63	0.05	0.25	-0	-0.2
8					1	0.39	-0	-0	0.55	0.35	0.01	-0	0.19	0.31	0.16	0.29	0.07	0.35	0.28	0.15	0.01	0.15	-0	-0	-0
10						1	0.02	-0.1	0.35	0.54	0.09	-0.2	0.13	0.03	0.08	0.04	0	0.57	0.22	0.07	0.3	0.18	0.05	0.09	-0.1
13							1	0.18	0.1	0.03	0.15	-0.1	-0	-0	0.04	-0	-0.1	-0.1	0.04	-0	-0.1	-0	0.02	-0.1	-0
14								1	0.18	0.27	0.06	0.03	-0.1	0.07	-0.1	-0.1	-0	-0.1	0.05	-0	-0.1	0.01	0.08	-0	-0.1
17									1	0.59	-0	0.03	0.01	0.35	0.05	0.02	0.03	0.18	0.24	0.04	-0.1	0.15	0.03	0.07	-0
19										1	-0	-0.1	-0	0.21	0.05	0.02	0	0.25	0.14	0.09	0.13	0.22	0.13	0.07	-0.2
22											1	-0.2	0.21	-0.2	-0.2	-0.1	-0.2	-0	0.08	-0.1	-0.1	-0.1	0.12	-0.1	0.16
23												1	-0.1	0.32	-0	0.15	0.06	-0.2	0.16	0.1	-0.2	-0	0.19	-0.1	-0
24													1	0.05	0.02	0.13	-0.1	0.18	-0	0.14	0.08	0.07	-0.1	-0	-0
27														1	-0	0.1	0.2	0.09	0.14	0.23	-0	0.05	0.34	-0.1	0.03
28															1	0.25	0.08	0.31	0.26	0.25	-0.1	0.12	-0.3	0.3	-0.1
29																1	0.08	0.29	0.17	0.27	0.01	0.04	-0.2	0	-0
31																	1	0.1	0.06	0.09	0.04	0.07	0.01	0.08	0.05
35																		1	0.47	0.44	0.18	0.15	-0.4	-0	-0
36																			1	0.32	-0.6	0.06	-0.2	0.02	0.2
37																				1	0.07	0.02	-0.2	-0	0.08
38																					1	0.1	0.11	-0	-0.2
39																						1	0.01	0.07	-0
50																							1	0.02	-0
59																								1	0.16
60																									1

Table 4.6. Uncorrelated features and their correlation values in the N-RevKent dataset

		Feature correlation values																								
		1	2	3	4	5	8	10	13	14	17	19	22	23	24	27	28	29	31	35	37	39	46	50	59	60
Feature correlation values	1	1	0.5	0.26	-0	0.21	0.59	0.45	0.1	0.04	0.34	0.24	0.26	-0.2	0.1	0.53	0.37	0.26	0.02	0.41	0.05	0.08	0.14	0.3	0.13	0.05
	2		1	0.32	-0.3	0.14	0.04	-0.3	-0	0.05	0.08	-0.2	0.19	-0	-0.1	0.21	0.08	-0.1	-0	-0.1	0.02	-0	0.21	0.44	0.1	0.16
	3			1	-0.1	0.12	-0.1	0.07	0	-0.1	-0.1	0.03	-0.1	0.09	-0	0.06	-0	-0	0.01	-0.2	0.01	0	0.07	0.31	0.11	-0
	4				1	0.07	0.25	0.11	0.08	0.06	0.22	0.07	-0.2	-0.1	0.1	0.39	-0.5	-0.3	0.08	0.04	-0	0.07	0.39	-0.1	-0.2	-0.1
	5					1	-0	0.19	-0	-0.1	-0.1	-0.1	-0.1	-0.2	-0	0.17	0	-0	-0	0	-0.5	0.08	0.13	0.22	-0	-0.2
	8						1	0.35	0.11	0.09	0.54	0.29	0.11	-0.1	0.13	0.31	0.15	0.19	0.01	0.33	0.05	0.06	0.17	-0	-0.1	-0
	10							1	0.06	-0	0.3	0.57	-0	-0	0.08	0.2	0.25	0.32	0	0.58	0.06	0.17	-0.2	-0	0.01	-0.1
	13								1	0.31	0.22	0.12	0.09	-0.1	0.05	0.14	0.09	-0	-0	0.03	0.01	0.05	0.1	0.05	0.07	0.01
	14									1	0.23	0.28	0.09	-0	-0.1	0.07	-0	-0	0.01	-0	0.05	0.06	0.13	0.12	-0	-0
	17										1	0.53	0.04	0.15	-0.1	0.28	0.13	0.1	-0	0.18	0.05	0.08	0.28	0.08	0.02	-0
	19											1	-0.1	0.08	-0.1	0.1	0.23	0.16	-0	0.29	0.09	0.18	-0	0.06	-0	-0.2
	22												1	-0.2	0.28	-0	0.12	-0.1	-0.1	-0.1	-0.1	-0.1	0.08	0.25	0.06	0.14
	23													1	-0.1	-0.2	0.03	0.1	0.13	-0.2	0.06	-0	0.07	-0	-0.2	-0
	24														1	0.09	0.07	-0	0.05	0.1	0.06	0.07	-0	0.01	0.03	0.1
	27															1	0.17	0.13	0.18	0.24	0.06	0.09	0.52	0.13	0.13	0.01
	28																1	0.29	0	0.3	0.07	0.07	-0.1	0.02	0.32	-0
	29																	1	0.13	0.42	0.23	-0	-0.1	-0.3	0.02	0
	31																		1	0.09	0.11	0.11	0.06	-0.1	0.04	0.02
	35																			1	0.45	0.1	-0.5	-0.5	-0	0.03
	37																				1	-0	-0.4	-0.3	-0.1	0.06
	39																					1	0.01	0.06	0.02	-0
	46																						1	0.41	0.02	-0
	50																							1	0.24	-0
	59																								1	0.25
	60																									1

Table 4.7. Uncorrelated selected features for O-RevKent and N-RevKent

Feature Number	**Feature Names**
1	Total distance of pen travelled
2	Total signature execution time
3	Pen lift:(Number of pen ups=> button 1 to 0)
4	Average velocity in X direction
5	Average velocity in Y direction
8	Maximum pen velocity in x - Average pen velocity in x
10	Maximum pen velocity in y - Average pen velocity in y
13	Average pen acceleration in x
14	Average pen acceleration in y
17	Maximum pen acceleration in x - Average pen acceleration in x
19	Maximum pen acceleration in y - Average pen acceleration in y
22	Azimuth
23	Altitude
24	Pressure
27	Standard deviation x coordinate values
28	Maximum x coordinate value - Last x coordinate value
29	First x coordinate value - minimum x coordinate value
31	Average x coordinate value
35	Standard deviation y coordinate values
37	First y coordinate value - minimum y coordinate value
39	Average y coordinate value
46	width/height ratio
50	Number of vertical midpoint crossing the signature
59	Average pen jerk in x
60	Average pen jerk in y

It is found from the obtained Spearman-rank correlation test results that features 1, 2, 4,5, 8, 10, , 13, 14, 17, 19, 22-24, 27-29, 31, 35-39, 50, 59, 60 are not correlated in O-RevKent while in N-RevKent, features 1-5, 8, 10, 13, 14, 17, 19, 22-24, 27-29, 31, 35, 37, 39, 46, 50, 59, 60 are found to be uncorrelated. In summary, feature 36 and 38 are found to be correlated in N-RevKent dataset while these features are not correlated with other features in O-RevKent. Features 3 and 46 are found to be uncorrelated in N-RevKent dataset while correlated with other features in O-RevKent. Looking at the definition as described in Table 2.1-Table 2.3 in Chapter 2, feature 36 and feature 38 are both based on the vertical (y) coordinate value and similar (but not correlated) to features 35, 37 and 39; where features 3 and 46 are very commonly used features in signature processing [116], [175], [209], [255], [256]. In order to ensure like-for-like analysis as far as possible in both O-RevKent and N-RevKent, features 1-5, 8, 10, 13, 14, 17, 19, 22-24, 27-29, 31, 35, 37, 39, 46, 50, 59, 60 will be utilised for both datasets in the experiments reported in this chapter. This feature subset is also shown in Table 4.7.

As described in Chapter 1, the handwritten signature is the product of a learned neurophysiological motor program which is a complex interaction of cognitive and neuromuscular and biomechanical [85], [86] processes. [85], [86], [191], [192], [257]. People generally start learning to write at a very early age, eventually develop a signature model of their own and practise constantly to produce similar signature samples according to their own personal model. But everyone is aware that even after practising constantly the appearances of two samples of a person's signature is not always the same. As well as the intra-person and inter-person variability described in Chapter 1, based on the effect over time the variability is also categorised as: short-term variability and long-term variability. Short-term variability is evident on a day-to-day basis as it depends on the signing condition such as the position of signing and pen grip as well as the writing instruments and writing surface used [91], [92], and the psychological or environmental (e.g. stress) [93], [94] condition of the signer at the time of signing, while long-term variability is apparent over longer periods of time as it depends on the modification of the parts of the motor system (e.g. brain, central and peripheral nervous systems, muscles, limbs etc.) [126], [191], [212], [258]. It may also be influenced by neurodegenerative diseases affecting the signer[96], [98], [187], [259]. Due to this variability and complexity in handwritten signatures, making an automatic signature verification system applicable in all everyday-life applications is still remains difficult. Several studies have been reported recently in [158], [251], [258], [260]–[263] have focused on the analysis of the signing process, and the variability and complexity of handwritten signatures in particular in order to achieve a better understanding of the complex phenomena underlying the signing process. Therefore, the adoption of the handwritten signature as a biometric modality depends on the reproducibility of signature samples in an individual. In other words, the "stability" in signing (the extent to which the "intrinsic properties of rapid human movements that constitute the basic element of each signature" [126] are reproduced) is a principal factor in determining the suitability of the signature for biometric identification.

To study natural revocability in signatures it is necessary to investigate the signing process in both the original and the new signature domains of individuals since, as with all behavioural biometrics, intrinsic variability within samples of any individual can be considerable, and the existence of so-called *goats* (those whose signatures quite naturally vary a great deal) [254]) is not uncommon. In assessing a new signature, the principal issue of interest here is to determine whether, and how quickly, the signing process attains a degree of stability in reproduction which makes its use as a biometric indicator viable. Clearly, such an assessment benefits also from knowledge of the stability properties of the original signature.

Although the concept of stability may be difficult to define formally or quantitatively, it is intuitive that, in this context, developing the habits of signing through repetition is important in establishing the automatic signing patterns required to decrease the dissimilarity between intra-individual signature samples and hence ensure the degree of reproducibility required for biometric identification. Thus, an informal and intuitive notion of increasing stability is adopted - referring to the tighter clustering of samples in multi-feature space, - which we expect to observe among the samples of a new signature as time passes, and which is a prerequisite in the biometrics context. Understanding how stability, described in this way, changes with time, is therefore a primary factor in assessing the value and viability of *natural revocability* as a practical strategy.

In order to observe the variations in stability in the original and new signatures across different capture sessions, initially Euclidean distances between samples in multidimensional feature space are measured. As mentioned earlier in this section the RevKent database is utilised for this experimental study and Euclidean distances are measured between samples of handwritten signature of an individual in each capture session for both original and new signatures for each user (e.g. distance between sample 1 and sample 2, sample 2 and sample 3 and so on for session 1(S1), and same way for session 2 (S2), session 3(S3) etc. for user 1, user 2 etc.). Then the mean of these distances (*D*) for each user for each session is calculated and the mean distance, *M,* for all users is calculated for each session. Comparing the distance of a signature sample to other signature samples of the same individual (signer) is used as a *posteriori approach* when evaluating the quality of a biometric sample and also used in signature stability considerations[264]. In this way, the lower the mean distance *M* is for one session, the more similar is the sample to the other samples of the that session, in other words samples are more "stable" within that session. Figure 4.1 shows the variation of both the original and new signatures in each of the capture sessions, measured as the mean distances between samples captured in first four successive sessions (as a majority of the subjects/users completed at least four sessions).

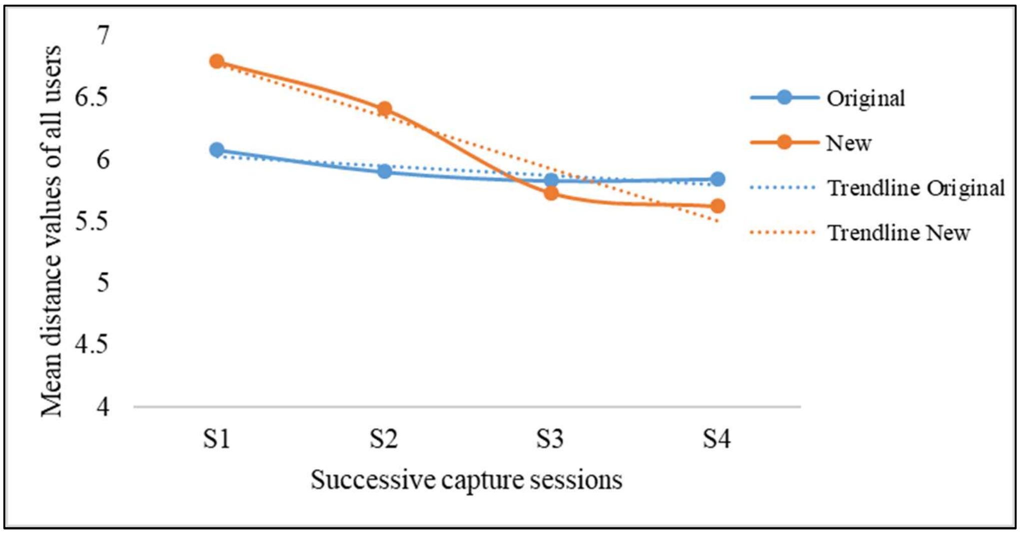


Figure 4.1. Mean distances between samples in each successive session (Session 1 to Session 4) for all users

It can be observed from Figure 4.1 that the mean distance between samples in the first session (S1) is higher than the second session (S2) and then the value of the mean distance in the second session (S2) is higher than the third (S3) and the fourth session (S4). Although it is likely that there is an effect here of unfamiliarity with the tablet-based writing environment, most likely to have an impact in the early session, it is clear from the trend line in Figure 4.1 that stability increases with time and that signature variability tends to stabilise as the sessions proceed. A similar analysis for an individual user's original and new signature is shown in Figure 4.2 (User 41 is shown in the figure which represents majority of the users except some '*goats*'). This also shows that the value of the distances between samples in the first session (S1) is higher than the other sessions and the second (S2), third (S3) and fourth session (S4) distance values gradually decrease, although the sample distances do not vary substantially across different sessions. But this nevertheless indicates that stability in a new signature can be achieved on a relatively short timescale.

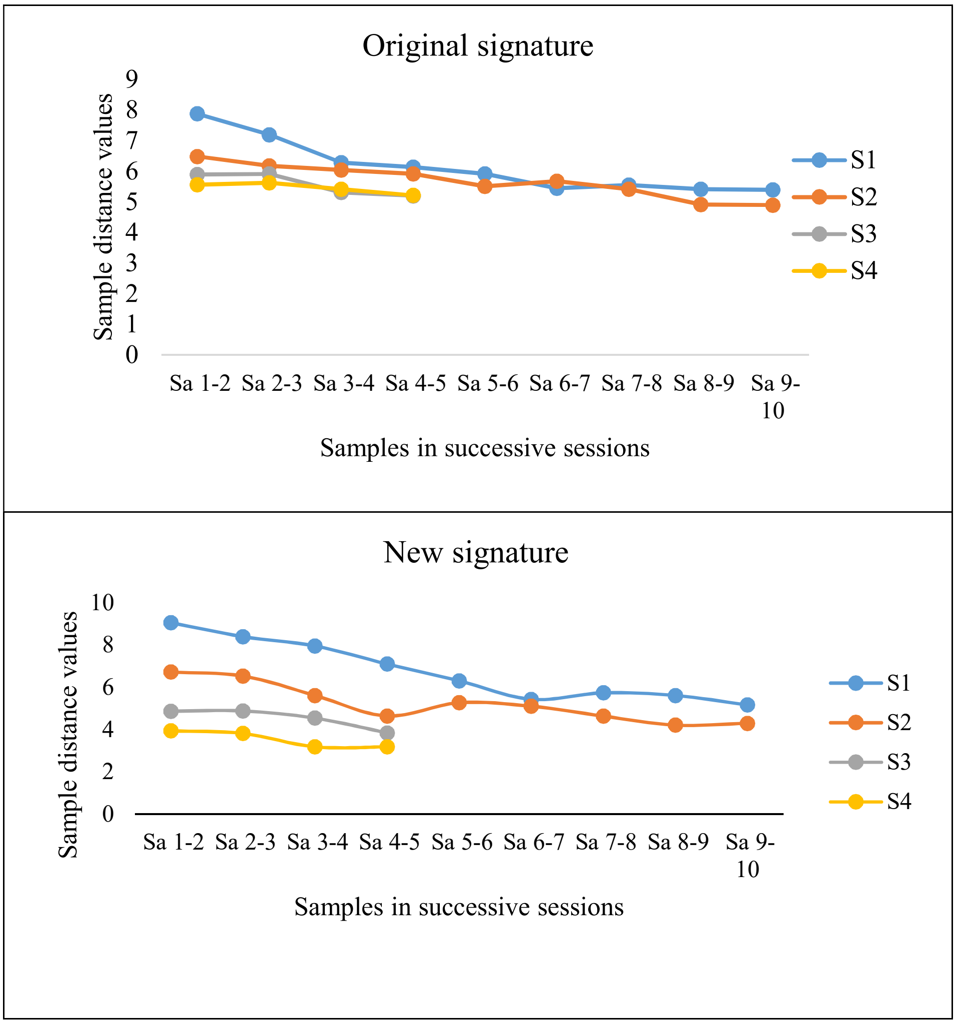


Figure 4.2. Distances between successive samples in each session for one user/subject (user 41)

As noted earlier in Chapter 3, it was challenging to recruit people for data collection, as participation was completely voluntary and motivated solely by a willingness to support the research effort without any monetary reward, and participants could withdraw anytime during the process between capture sessions. This made it especially difficult to collect data over longer periods (larger number of sessions) from the same participants. However, some participants contributed their signature samples for up to six and some up to ten sessions beyond the initial four collection sessions.

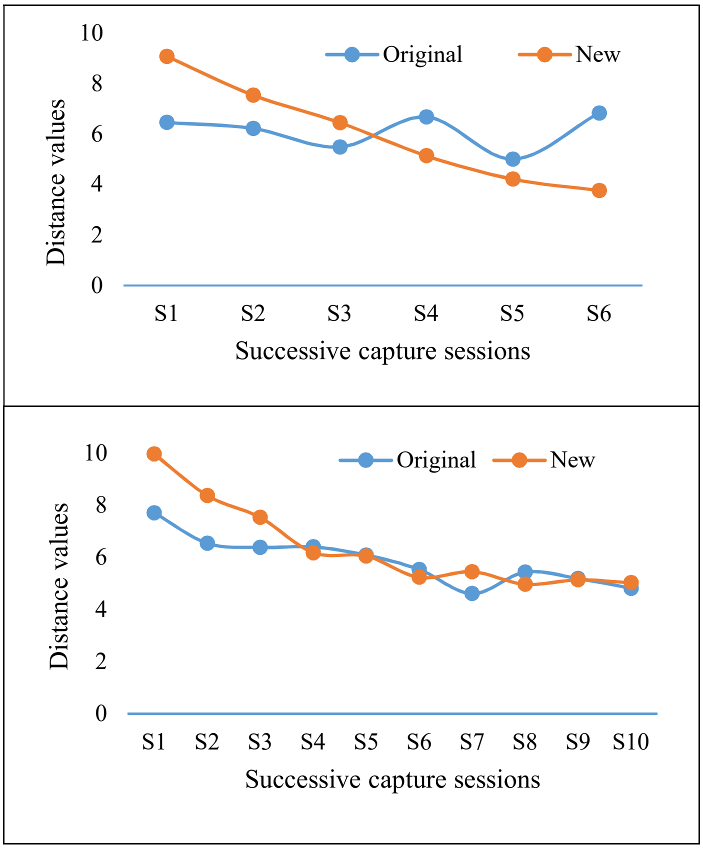


Figure 4.3. Distances between successive samples in each session for six (top) and ten (bottom) sessions for two different users

Figure 4.3 shows examples of two users' original and new signatures over this greater number of sessions, here six and ten acquisition sessions respectively, and Figure 4.4 shows the variation of both the original and new signatures in each of the capture sessions, measured as the mean distances between samples captured in ten successive sessions.

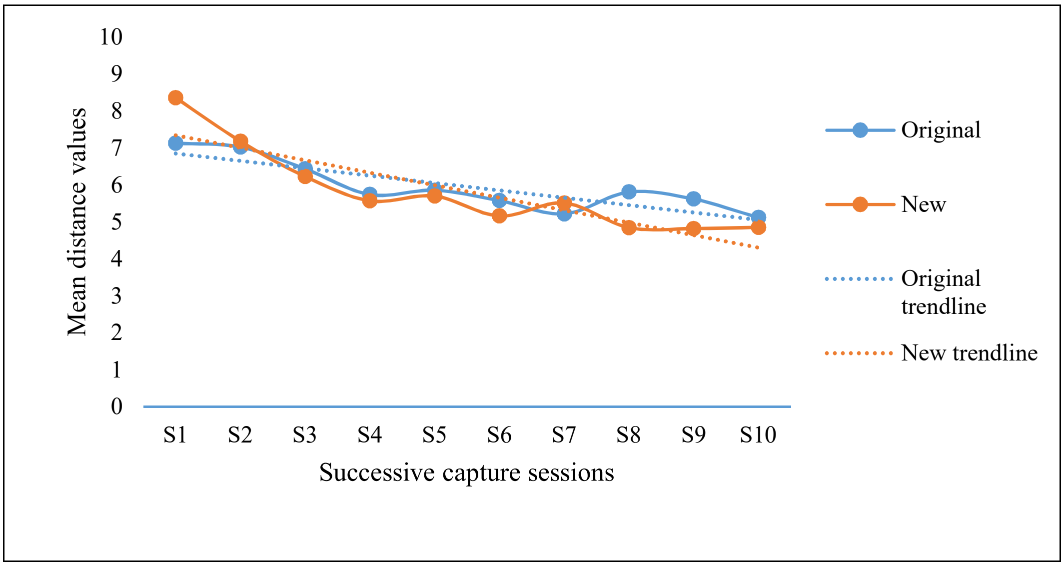


Figure 4.4. Mean distances between successive samples in each session for ten sessions (considering all users)

To get a more generalised and quantitative view of this increasing stability in the new signature a 'degrees of stability – low, medium and high' (categories defined in [261], [262]) analysis is performed for both signatures of all signers. For this the maximum and the minimum value of the mean distances of the original signatures are calculated and then two different distance values (T1 and T2) are set as threshold for the three degrees of stability (low, medium and high) as defined in (4.1) and (4.2) for the degree of stability.

$$T1 = M_{min} + T \tag{4.1}$$

$$T2 = T1 + T \tag{4.2}$$

where T = (M_{max} –M_{min}) / 3; M_{max} = Maximum of all distances M's of original signature; and M_{min} = Minimum of all distances M's of original signature. Using these threshold values, stability analysis is performed for each signer according to the degrees of stability defined in Table 4.8.

Table 4.8. Degree of stability

Distance values	**Degrees of stability**
Distance M >T2	Low Stability
T1< Distance M < = T2	Medium Stability
Distance M < T1	High Stability

Table 4.9. Percentage of achieved degree of stability in four successive sessions

	Degree of stability	**Sessions**			
		S1	**S2**	**S3**	**S4**
Original	High Stability	15.38%	25.64%	53.85%	61.54%
	Medium Stability	79.49%	71.79%	38.46%	33.33%
	Low Stability	5.13%	2.56%	7.69%	5.13%
	Total	100%	100%	100%	100%
New	High Stability	25.64%	28.21%	43.59%	53.85%
	Medium Stability	53.85%	66.67%	56.41%	46.15%
	Low Stability	20.51%	5.13%	0.00%	0.00%
	Total	100%	100%	100%	100%

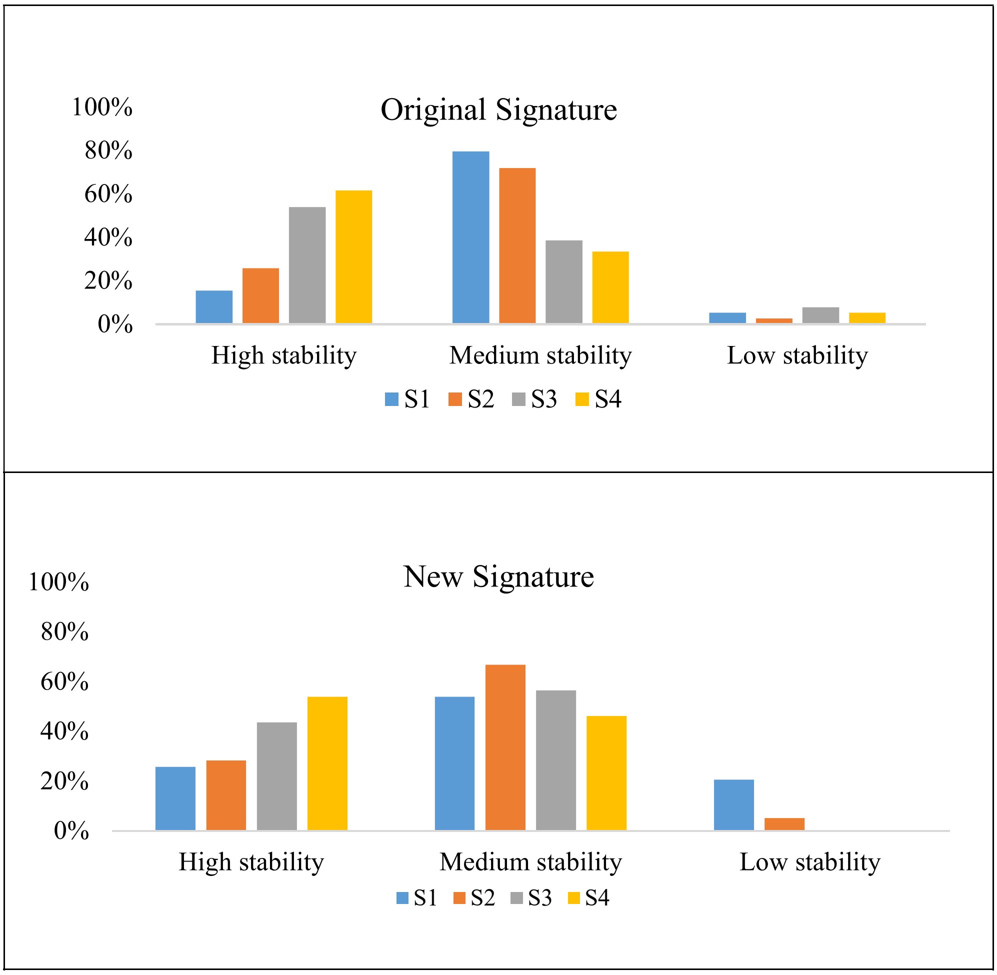


Figure 4.5. Degree of stability in each session for original and new signature (four sessions)

Table 4.9 and Figure 4.5 report the degree of stability found in each session in both original and new signatures, where all signers are considered. It is noticeable that comparing the original signatures and the new signatures (Table 4.9 and Figure 4.5), the percentage of high stability samples among the new signatures is higher in session one (S1) than that of the original signature. This may reflect a greater degree of care and caution on the part of the signer when developing a new signing style but, more importantly, it shows that the high stability increases and low stability decreases with time and also that the majority of the signers show either high or medium stability when signing the new signatures.

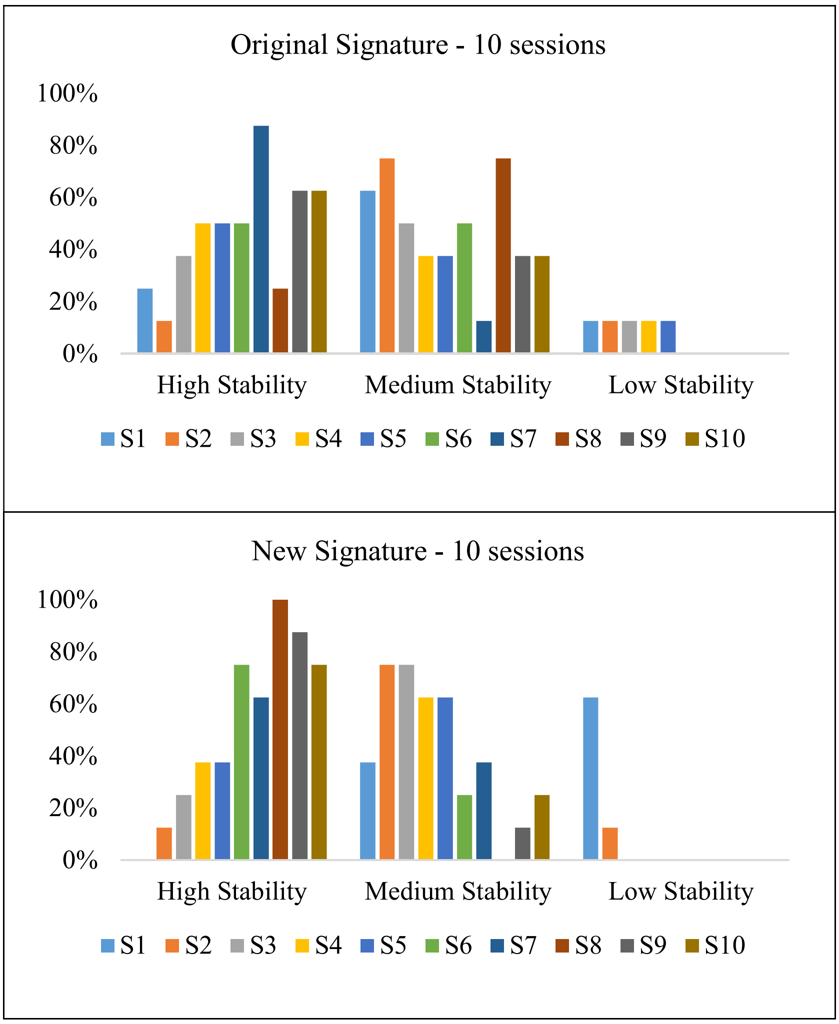


Figure 4.6. Degree of stability in each session for original and new signature (ten sessions)

Figure 4.6 reports the degree of stability analysed in the same way for a longer number of capture sessions, in this case ten sessions but for the smaller number of signers who had volunteered to contribute both the original and new signatures for ten capture sessions. While it is important to exercise some caution when dealing with such a small number of signers, these results do suggest, on the one hand, that behavioural biometrics are always open to the possibility of somewhat unpredictable characteristics, but also that if a sufficient time period is allowed then there is a possibility of convergence in stability between a highly practised and long-standing signature and an alternative new representation.

4.3.1 Categorised analysis

In order to investigate the characteristics of potential revocability in the signature modality, it is useful to analyse performance by invoking the "biometric menagerie" notation for individual behaviour first introduced by Doddington et al. in the context of speaker recognition [254]. The authors in that study grouped speakers in four categories, labelled respectively "*Sheep*", "*Goats*", "*Lambs*" and "*Wolves*" (Figure 4.7) based on the given classifier's genuine and non-genuine match scores.

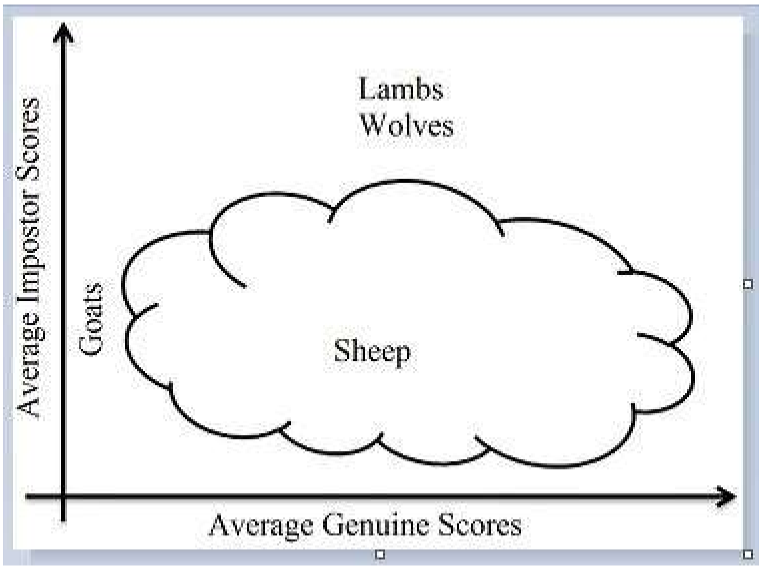


Figure 4.7. Doddington's representation of the "biometric menagerie" [48]

According to [254] the speakers who are recognised easily are labelled as "*Sheep*"; speakers who are difficult to recognise are labelled as *"Goats"*, speakers who can be imitated easily are labelled as *"Lambs"* and "Wolves" are the speakers who are good imitators (i.e. they are particularly successful at imitating others). This "biometric menagerie" notation has been later applied in the context of other biometric modalities such as, fingerprint [265]–[267], iris and face [265], [267], [268]. More categories of users (*"Worms", "Chameleons", "Phantoms"* and *"Doves"*) have been added recently by Yager & Dunstone [76], [78] to the "biometric menagerie". All these study focusing on Doddington's categorisation [46], [265], [267], [268], [73]–[75] provide evidence that the "biometric menagerie" can be a useful concept in most biometric modalities.

In the present study we are especially interested in characterising individuals as sheep or goats, designated according to the following definitions:

- **Sheep**: Sheep, in this model, are those signers who show relatively little variability in their signature samples over time (i.e. those whose signatures are generally stable).
- **Goats**: Goats are those signers whose signature samples have a tendency to considerable variation over time (i.e. those whose signatures are generally unstable).

In this way both the original and newly invented signatures have been observed to determine the extent to which the characterisation of an individual's signing behaviour remains constant between the original and new signature style, or whether and how individuals change category. To this end an analysis of each individual's signatures was carried out with respect to their signature stability category, with results summarised in Figures 4.8, 4.9, 4.10, and 4.11. Here we use a rather subjective and intuitive interpretation of "stability" for the purpose of a qualitative analysis. For example, if the mean distances measured between signature samples captured in successive sessions gradually decrease or do not vary significantly with time then this signature can be deemed stable in this context. Figures 4.8, 4.9, 4.10, and 4.11 show the different groups of signers within the test population categorised according to the relationship between the original and new signing characteristics.

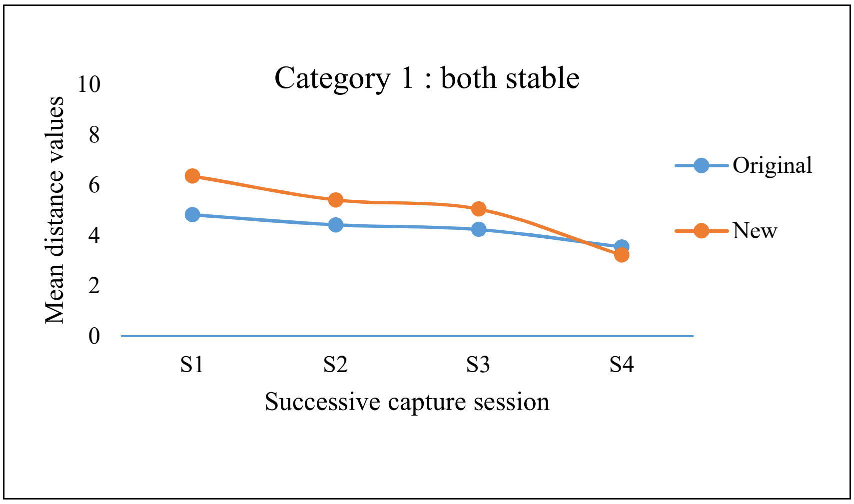


Figure 4.8. Category 1- Signers' both original and new signatures are stable (Consistently sheep)

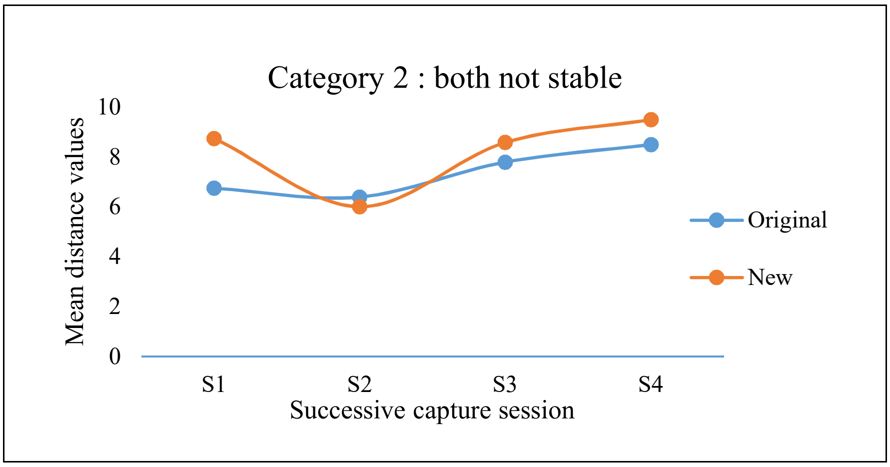


Figure 4.9. Category 2- Signers' both original and new signatures are unstable (Consistently goats)

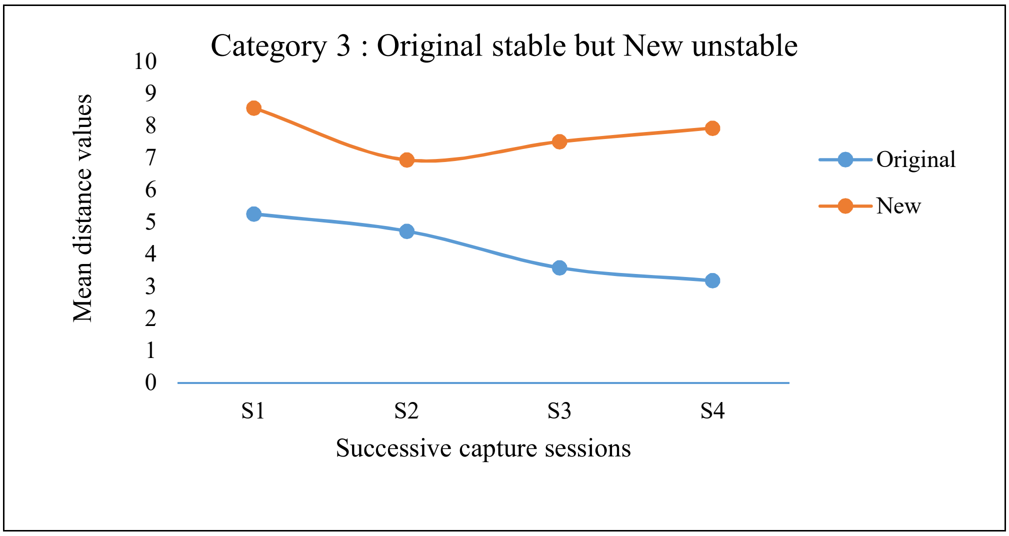


Figure 4.10. Category 3- Signers' original signature is stable but the new signatures is unstable (Sheep change to goats)

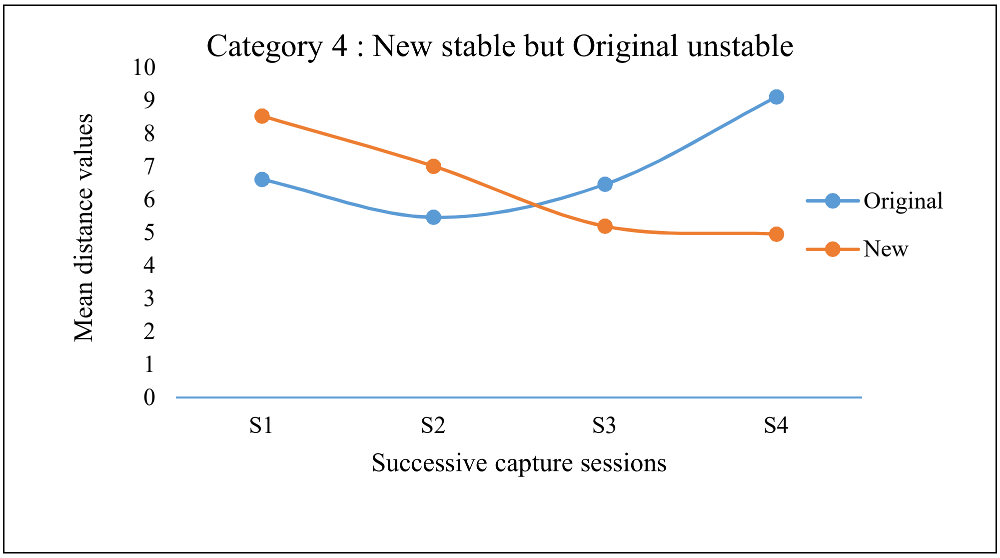


Figure 4.11. Category 4- Signers' original signatures are unstable but new signatures are stable (Goats change to sheep)

The signer population can then be divided into four categories, as follows:

- Category 1- In our current context, this category defines those individuals who are consistently sheep according to our qualitative definition (i.e. those for whom both the original and new signatures are stable - Figure 4.8).

- Category 2 - These are those signers who are consistently goats (i.e. those for whom both original and new signatures are unstable - Figure 4.9).

- Category 3 – This category defines those signers who were sheep with respect to their original signatures, but who turned into goats when generating a new signature form (i.e. those whose original signatures are stable but whose new signatures are unstable - Figure 4.10).

- Category 4 – These are signers who were goats with respect to their original signature samples, but turned into sheep when developing a new signature (i.e. whose original signatures are unstable but whose changed new signatures are stable - Figure 4.11, this could be due to paying more attention when creating the new signature or creating a different signature style or less complex signature than the original signature which makes the new signature less variable).

To get an estimate of the existence of these signers' categories within the overall test population a further analysis is performed based on the qualitative definition of the sheep and goats as shown in Figure 4.12.

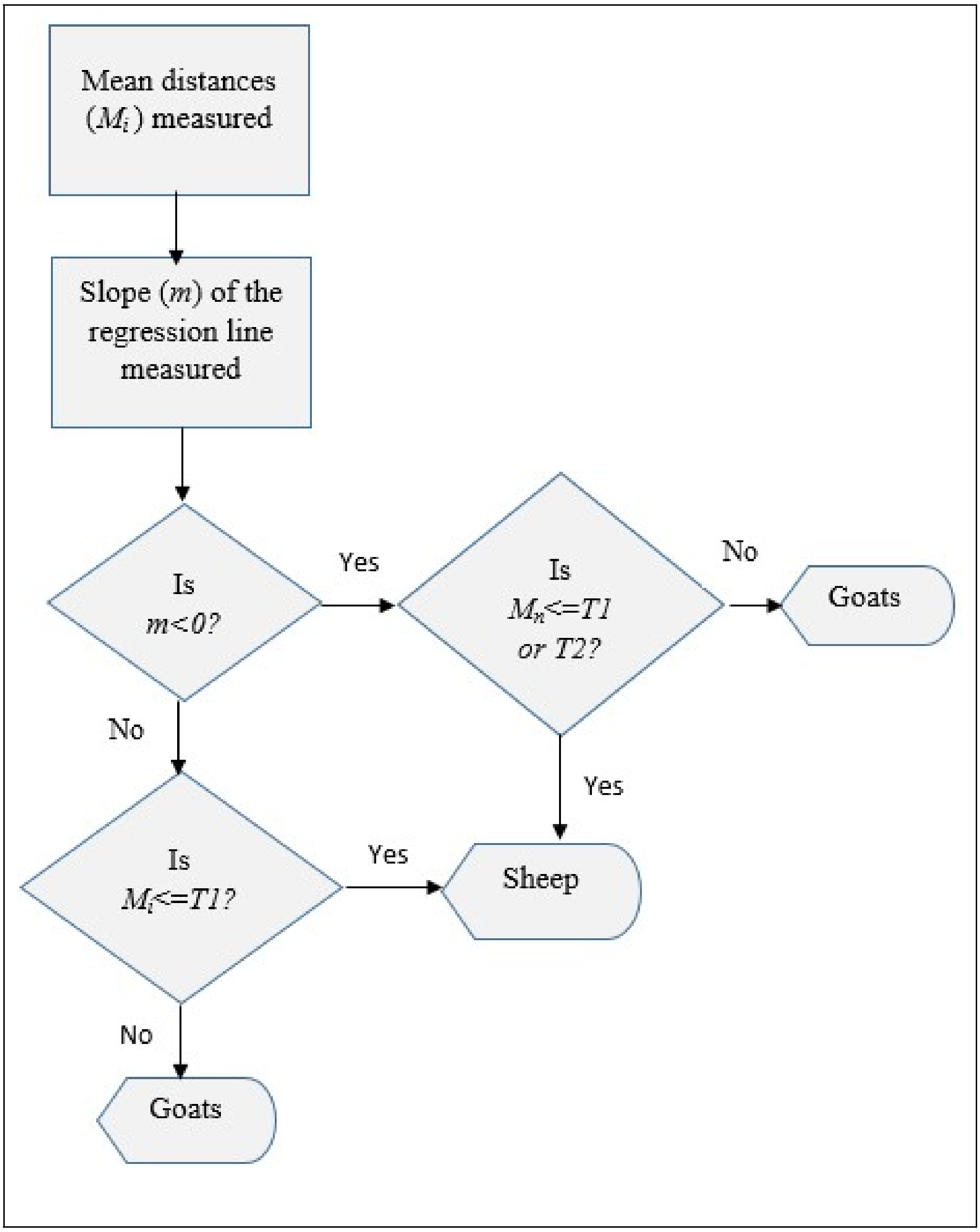


Figure 4.12. Defining sheep and goats.

To measure the increasing or decreasing pattern of the mean distances (M's) for the successive capture sessions a linear regression model is developed and the slope of the regression line is measured for both the original and the new signature for each signer as defined in (4.3).

$$\boldsymbol{Slope\ (m\ )} = \frac{\sum_{i=1}^{n} (S_i - \bar{S})(M_i - \bar{M})}{\left(\sum_{i=1}^{n} (S_i - \bar{S})^2\right)} \tag{4.3}$$

where i = 1, 2, 3… n, and n is the number of capture sessions, S_i and M_i are the capture sessions and the mean distances respectively for the i^{th} session, and $\bar{S}$ and $\bar{M}$ are the mean of S_i and M_i respectively. Each value of m obtained by the evaluation is either positive (above zero) or negative (below zero), and a positive value of m represents an increase in mean distances M with time while a negative value of m represents a decrease in mean distances M with time, which is an indicator of stability of the signature with time. As the handwritten signature is a behavioural biometric, and due to its natural variability, there may be some cases where the mean distance is very low from the first capture session but shows a very low positive slope and the mean distance is low for all the sessions (i.e. though the distances do not show a decreasing pattern they are in the low distance (High stability, M<=T1) range for all the sessions (Figure 4.13). On the other hand, there may be cases where the mean distances show a decreasing pattern with a low negative slope but the value of the mean distances are in the low stability range (M>T2). In the latter case, this is showing a decreasing pattern which means it could reach a high stability range if more time were allowed, but for the purpose of this analysis this type of signature will be deemed as unstable (goats) due to having four capture sessions (as majority of the signers' completed up to four sessions, therefore, session one, two, three and four are chosen for a like-for-like analysis across all the signers). Taking these cases into consideration, sheep and goats are determined for all the signers as shown in Figure 4.12 and the percentage of each category of signers (as described earlier in this section) within the overall population is shown in Figure 4.14.

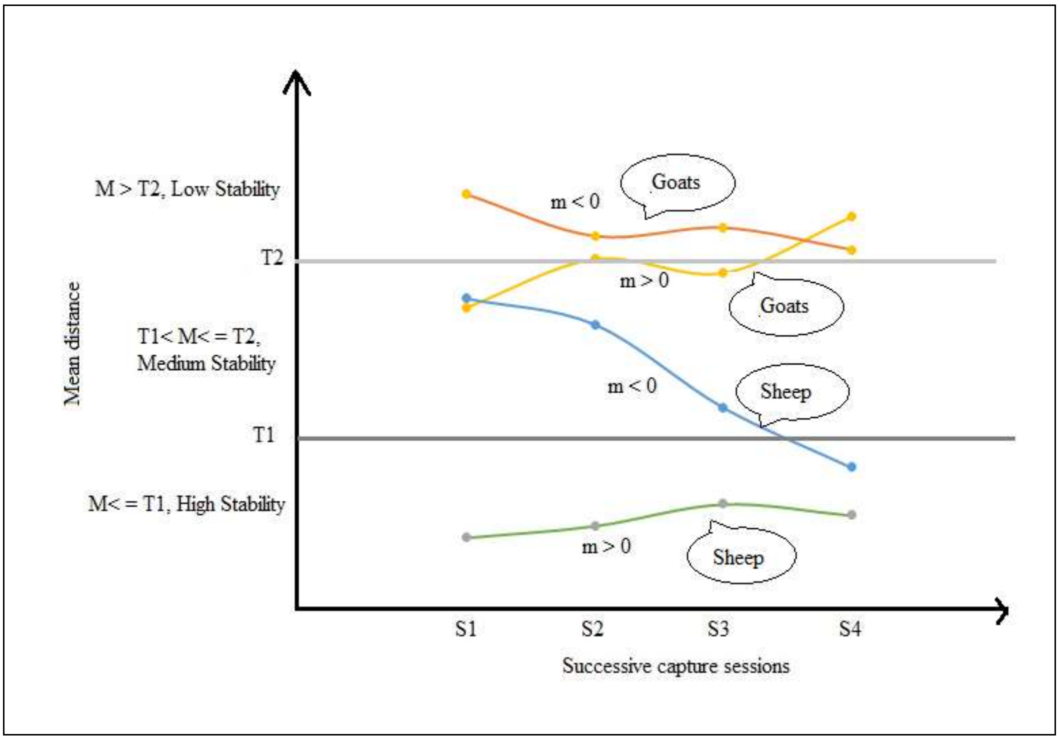


Figure 4.13. Examples of different cases of signers' stability

Figure 4.14 shows that Category 1 contains the largest number of individuals across the population, indicating that stable signers in their original signatures remain stable signers when developing a new signature. Encouragingly, a rather small proportion of the population are unstable signers in their original signature and remain so when moving to a new signature style, but some 10.26% of the population with an originally stable signature generate a degree of instability when changing to a new signature style - although we have presented evidence that such a group may, in the longer term, still achieve stability. Finally, for 15.38% of the population who exhibited instability in the original signature the results show that it is possible nevertheless to achieve stability when changing their signature style.

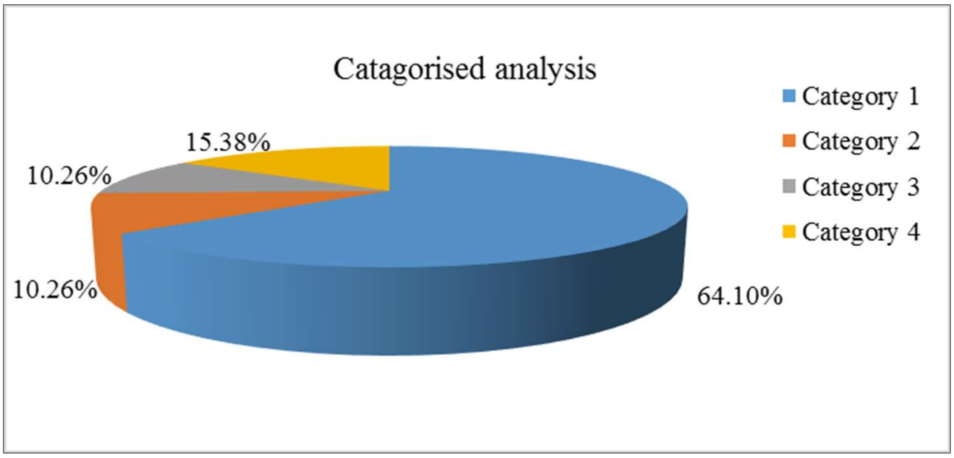


Figure 4.14. Percentages of each category in overall population

4.3.2 Classification results

A more practically-oriented test of the viability of the natural revocability concept for the handwritten signature may be considered to be the achievable performance directly in a recognition/verification scenario. To observe whether the newly formed signature can be reliably verified or not, and to compare the performance achieved against performance with the original signature, a recognition experiment was performed using the data mining software named Weka [220]. Since for our purposes in the present context, relative performance is more important than absolute performance, for this experiment a simple nearest neighbor classification algorithm (KNN) [219] was adopted. However, in order to give greater generality to the results, performance using the SVM classifier [272] was also determined.

As described in Section 2.5 in Chapter 2, KNN is a simple, non-parametric classifier that does not require any explicit training phase but the class or category representing the samples must be labelled. In these experiments the simplest configuration with k = 1 has been used. The first nearest neighbour is determined by measuring the squared Euclidean distance between the test signature sample and all the training signature samples. Then the measured distances are sorted from minimum to maximum and the class label appearing within the minimum distance (first nearest neighbour) is assigned to the test sample.

The Support Vector Machine classifier approach was originally designed for binary classification. It finds a hyperplane or a decision surface (separating plane) between two point classes determined by certain points (the closest points between the two classes) of the training set. The closest points are called the support vectors. The hyperplane or the decision surface tends to maximise the distance. The training samples are divided into positive and negative groups according to this hyperplane or surface. In this experiment, the 'one-against-one' method reported in [273], has been adopted in order to apply the binary SVM classifier to a multi-class problem. The method constructs N(N – 1)/2 binary classifiers (N being the number of classes) for each distinct pair of classes and then each binary classifier assigns the test sample to one of the two distinct classes . In order to assign a class label to the test sample, a majority voting strategy is used. In each case, the assigned class label is incremented by one and finally the class with the most votes determines the ultimate class label of the test sample. In the case of ties, the first class label is assigned to the test sample.

For both the KNN and SVM classifiers a 10-fold cross validation methodology [274] has been adopted, where the dataset is randomly divided into ten subsets of approximately equal size. One of these ten subsets is used as a test set and the remaining nine subsets are used for training the classifier. This process is repeated ten times until each of ten subsets is used as a test subset once. If the class of the predicted training and test subjects match, are from the same class, the decision is recorded as correctly classified; otherwise it is recorded as incorrectly classified. The overall error rate is evaluated as the percentage of total number of incorrectly classified test subjects out of all the test subjects. Figure 4.15 shows the evaluated error rates for the original and new signatures using both KNN and SVM classifiers. The classification results are based on one-to-many identification and recognition rate is on the basis of the correct match in the first rank (Rank 1). The reported results show that the error rate generated is lower in both classifiers for the new signature than for the original signatures, although the difference is modest. This may be due to taking greater care when creating the new signature, or the style or type of the new signature which may have more distinguishing features than the original signature for recognition (signature style and a feature based analysis is reported in Chapter 5). But, nevertheless, this shows the potential of the new signatures to be reliability recognised in the event of compromise of the original signature without degrading the recognition accuracy.

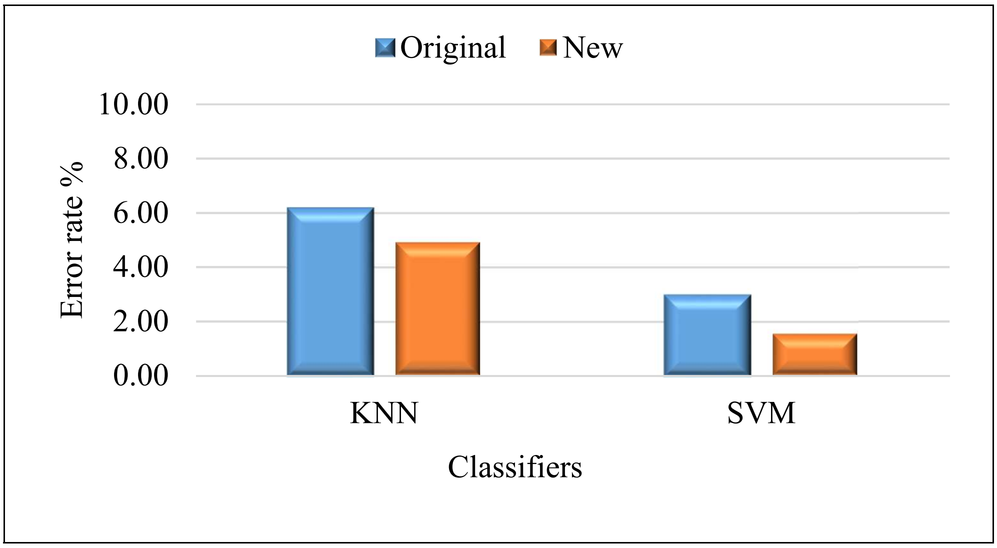


Figure 4.15. Classification results

4.4 Conclusions

In this chapter, the idea of natural revocability – a particular characteristic of a behavioural modality which might be exploited in security applications, has been investigated and explored for the handwritten signature modality. Experimental analysis has been performed using both O-RevKent (Original signatures) dataset and N-RevKent (New signatures) dataset, which enabled the investigation to carry out using the current familiar signing pattern of each individual as a reference point.
.

Although it is difficult to assess the wider significance of the results using only a small database, the observations provide initial indications which suggest, in general, that individuals are able in most cases to develop a new signature which can quickly achieve stability. This provides some evidence, and a degree of optimism, that the idea of natural revocability, whereby a compromised signature can be revoked by a user and a new signature form substituted, offers potential viability in a practical scenario, especially given that stability appears in most cases to be achieved over a relatively short timescale. It has also been shown how stability patterns between original and new signatures change across individuals. Finally, recognition performance has been evaluated as a more practically-oriented test of the viability of the natural revocability concept. It has been shown that the new signatures can be reliably recognised and even that the performance achieved can be better than that found for the original signatures. Further and much more extensive investigation would be required to determine how this might change over a longer timescale, such as applies when the original signature, typically developed and established over long periods, is used.

In summary, the study presented in this chapter provides some valuable indicators of performance in respect of the potential for further development of the principles of natural revocability as a practical option in appropriate applications.

The next chapter will explore some further issues related to the concept of natural revocability such as correlation between original and new signatures, relationship between performance and the features used and so on.

Chapter 5:

Feature Based Analysis of Natural Revocability

This chapter will present some feature-based analysis of the natural revocability concept (defined and presented in Chapter 4) by investigating some features commonly used in signature processing in both original and new signatures of a group of writers, and exploring the relationship between features, signature style and their effect in relation to original signatures and new signature. Section 5.1 will give a brief review of related feature-based studies. Section 5.2 will investigate the relationship between original and new signatures based on feature values, variability and interrelations between them. Following this some experimental analysis will be reported in Section 5.3; examining the effect of signature style in sub-section 5.3.1 and the effect of demographic factors (specifically, age, gender and handedness) in sub-section 5.3.2. Finally, Section 5.4 will conclude this chapter.

5.1 Review of feature based analysis in handwritten signatures

Since the handwritten signature is a very widely used biometric trait (discussed in Chapter 1), a considerable number of studies have been reported on some particularly important factors related to signature verification and identity prediction, for example, reliable signature verification systems [17], [198], [255], [275], different types of approaches to extract information from signature samples [126], [276], [277], the investigation of the consistency of signatures [210], [278], improving identity prediction performance using soft biometrics [56], investigation of ageing in signature biometrics [153], [213], [279] and so on. Most studies within the signature biometric field have examined both the static data relating to the physical characteristics of the signature and the dynamic data relating to often more informative [93], [280] temporal characteristics of the signing behaviour (in other words how the signature was constructed). Some feature based analyses have also been reported, including, analysis of feature repeatability or stability [207], where a set of common, mostly static features have been used to analyse feature variability both within a single capture session and over time (multiple sessions) and the influence of physical characteristics on the variability.

In another study reported in [200], the signing process is analysed using dynamic features covering both the kinematic (describing the motion) and kinetic (describing the force involved in the movement) characteristics of the signing behaviour in order to provide evidence of the distinctiveness of genuine and forged handwritten signatures. Differences and similarities of the dynamic features (pen position, writing velocity, pen stop or break, pen force, pen lifts and landing, pen orientation - azimuth and pen tilt) between genuine and forged signatures have been reported.

A recent similar feature-based study has been reported in [281], which performed dynamic signature verification under a forensic scenario using a large number (40 out of 117 were selected for verification) of global features to improve performance. This study also reported some statistical analysis of a set of selected features (6) for genuine and forged signatures. A set of selected features have been analysed statistically for genuine and forged signatures to obtain useful information that could support forensic experts in distinguishing between genuine and forged signatures.

Another study in the forensic field, reported in [214], examined whether there is any difference in dynamic features (duration, size, velocity, jerk, and pen pressure) between genuine and simulated signatures and if the signing style (text-based, stylized, and mixed) has any impact on the characteristics of handwriting movements for simulations. The author also reported another study [282] where the differences in the same dynamic features between different signature styles have been examined. The effects of signing style on handwriting dynamics across genuine, disguised and auto-simulated signatures have also been examined. The results reported in both studies suggest that the signing style might be a potentially significant characteristic in signature feature evaluation with a view to forming opinions regarding authenticity.

The impact of signature legibility [283] (legible, medium and non-legible) and signature type (Simple flourish, Complex flourish, Name + simple flourish, and Name + complex flourish) in off-line signature verification and classification of handwritten signatures based on name legibility [177] have also been reported in the biometrics field. Some other feature based studies reported in the literature include the relationship between personality trait and signature production [284], feature based comparison of pen-based and swipe-based signature characteristics [285].

5.2 Relationship between features in original and newly established signatures

This section will explore the relationship between writers' original and newly-established signatures, initially based directly on feature values, and later considering the variability of features. The effect of feature values on feature variability is also investigated, observing the differences and similarities of the effect between original and new signatures in the same writer.

For this experimental study, 60 commonly used [153], [177], [207], [215], [255], [282], [284], [286] features, as defined in Table 2.1, 2.2 and 2.3 (in Chapter 2), are extracted from all the signature samples of the two datasets O-RevKent (original signatures) and N-RevKent (new signatures) of the Rev-Kent database (the data collection and dataset characteristics are described in detail in Chapter 3). Following feature extraction, a Spearman rank correlation test (as defined in Chapter 2) is performed between each corresponding feature in the original and new signatures. As described in Chapter 2 (Section 2.4) the obtained correlation coefficient *rho(ρ)* is a number between –1 and 1 that determines the extent to which the feature values in the original and new signatures are related. A *rho(ρ)* value close to zero indicates there is no evidence of any correlation or relationship, while the closer to 1 is this value, the stronger is a positive correlation (i.e. if the feature value in the original signature increases, the feature value in the new signature also increases) while the closer to –1 the stronger is the negative correlation (i.e. if the feature value in the original signature decreases, the feature value in the new signature also increases). A significance level (p-value) is also calculated to determine the confidence in the relationship (as described in [287]) between the feature values in the original and new signatures. The *rho(ρ)* value indicates the strength of the correlation between original and the new signature for a feature (e.g. according to [288] strength of the correlation based on *rho* (*ρ)* value: 0.9-1 very strong, 0.7-0.9 strong, 0.5-0.7 moderate, 0.3-0.5 weak and 0-0.3 negligible) and the *p-value* indicates how significant the relationship or the correlation is between original and new signature for that feature. For example, if the calculated correlation coefficient and p-value for feature 24 between original and new signature is 0.91 and 0.001 respectively, this indicates that feature 24 in original and new signature is very strongly correlated and correlation is highly significant i.e. there is over 99% chance that feature 24 in original an new signature is very strongly correlated. The results of the correlation (correlation coefficient *rho* (*ρ)* and its significance level *p-value)* are shown in Table 5.1.

It can be observed in Table 5.1 that many features are found to have significant positive correlations between the original and new signatures (with some features having very strong correlations and some having moderate to weak correlations). For example, azimuth (feature number 22), altitude (23), pen pressure (24), horizontal centralness (42), average y coordinate value (39) are found to have strong correlations between the original and new signatures. Average resultant velocity and acceleration (53 and 58 respectively), difference between maximum and average and between maximum and minimum vertical acceleration (19, 20) are found to have moderately strong correlations. This indicates that the same underlying constructional mechanism is evident in the new signatures as was the case with the original signatures for these features. It is also noted from Table 5.1 that there is no evidence of significant correlation found in total distance travelled (1), signature execution time (2) number of pen lifts (3), signature width (44) between original and new signatures. Figure 5.1 shows the boxplots of feature values of these features in the original and new signatures. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points considered.

Table 5.1. Correlation between features in the original and new signatures (NS = Not significant, considered if p-value >0.05)

Feature No	*rho(ρ)*	*p-value*		Feature No	*rho(ρ)*	*p-value*
1	0.177126	NS		**31**	0.5731	p<0.001
2	0.290688	NS		**32**	0.3453	p<0.05
3	0.198623	NS		**33**	0.1186	NS
4	0.592105	p<0.001		**34**	0.4698	p<0.01
5	0.687854	p<0.001		**35**	0.5393	p<0.001
6	0.606478	p<0.001		**36**	0.6561	p<0.001
7	0.574089	p<0.001		**37**	0.7024	p<0.001
8	0.593725	p<0.001		**38**	0.6182	p<0.001
9	0.489271	p<0.01		**39**	0.9279	p<0.001
10	0.768219	p<0.001		**40**	0.7802	p<0.001
11	0.799798	p<0.001		**41**	0.4759	p<0.01
12	0.677733	p<0.001		**42**	0.9279	p<0.001
13	0.339271	p<0.05		**43**	0.5731	p<0.001
14	0.274494	NS		**44**	0.2435	NS
15	0.685223	p<0.001		**45**	0.5852	p<0.001
16	0.651012	p<0.001		**46**	0.3686	p<0.05
17	0.718421	p<0.001		**47**	0.5034	p<0.01
18	0.741296	p<0.001		**48**	0.6321	p<0.001
19	0.838057	p<0.001		**49**	0.2365	NS
20	0.837854	p<0.001		**50**	0.4196	p<0.01
21	0.802429	p<0.001		**51**	0.6721	p<0.001
22	0.867409	p<0.001		**52**	0.7172	p<0.001
23	0.871862	p<0.001		**53**	0.7249	p<0.001
24	0.912348	p<0.001		**54**	0.7123	p<0.001
25	0.446761	p<0.01		**55**	0.6204	p<0.001
26	0.382794	p<0.05		**56**	0.3112	NS
27	0.293927	NS		**57**	0.5126	p<0.001
28	0.502632	p<0.01		**58**	0.8097	p<0.001
29	0.510121	p<0.01		**59**	0.6107	p<0.001
30	0.488462	p<0.01		**60**	0.5121	p<0.01

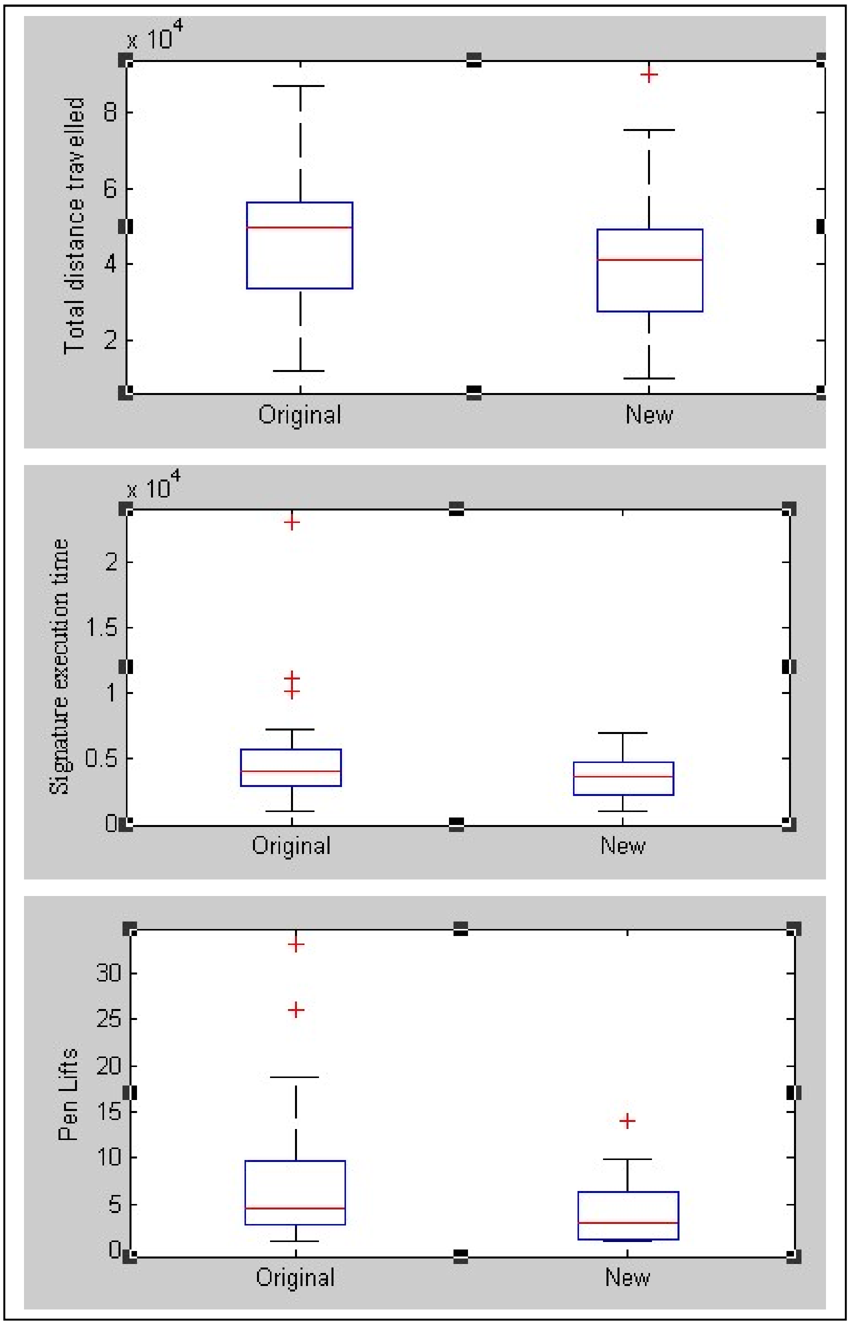


Figure 5.1. Boxplots of total distance travelled by pen, signature execution time, pen lifts in original and new signatures

It can be observed in Figure 5.1 that users tended to lift the pen more when writing their original signatures compared to when writing the newly formed signatures. Total distance travelled by the pen, and signature execution time are also slightly longer in the original compared to the same features in the new (naturally revoked) signatures. Whether the differences in feature values are due to creating a different signature or whether any other factors such as signature style and/or user/subjects age or gender has any effect on this, needs further investigation, which is discussed later in this chapter (in Section 5.3).

As described in Chapter 4, the physical dimensions of a handwritten signature can vary over time [211], [249]–[251] due to its natural variability, signing environment and so on, but the fundamental characteristics of the handwritten signature remain relatively constant over a period when written in a given frame. Therefore, it is particularly important to understand the variability of features of the new signature (for example, which features are more variable and which are more consistent) and comparing the variability with that of the original to see if the variability of the features is similar or different. In order to investigate this, features are extracted from the two datasets specified earlier in this section. Then, the mean, μ, and standard deviation, σ, of the feature values are extracted from the same and different acquisition sessions (intra-session and inter-session values respectively) for each feature for each user or user. To determine the variability of each feature the coefficient of variation (CV) is used – a measure of the dispersion of data that describes the variability relative to the mean [289]. It is measured as the ratio of the standard deviation, σ, to the mean, μ, as defined in (5.1). As this is dimensionless, it permits the comparison in variation of different feature values free from scale effects.

$$CV = \frac{\sigma}{|\mu|} \times 100 \tag{5.1}$$

In this way, an intra-session coefficient of variation (intra-CV) is calculated for each feature using the samples collected within a session for each user. This intra-CV for each feature describes how much the feature varies within the session - in other words, how repeatable, reproducible or stable the feature is in that session. Hence, a high intra-CV for a feature indicates that high variation occurs between the signature samples within a session and the feature is not repeatable or not stable (i.e. extracted feature values differ greatly between samples) whereas a low intra-CV indicates the feature is stable (i.e. extracted feature values are similar between samples). Also, to observe the features which show stability within a session, whether these show similar stability over a period or not, an inter-session coefficient of variation (inter-CV) is calculated for each feature using all the samples collected from all user's capture sessions. Like the intra-CV, a low inter-CV for a feature indicates that it does not vary greatly between all the sessions - in other words, the feature shows stability over a period for a user. Figures 5.2 and 5.3 show the boxplots of each feature's intra and inter session variability in the original and new signatures. The boxplots of each feature show the mean and the variance of the CVs across the users. It can be observed from these Figures that individual feature's variability is similar in the original and the new signature.

A further analysis has been carried out to obtain a generalised view of each feature's variability (or stability) from these intra-CVs and inter-CVs across all the users. A mean intra-CV (M-intra-CV) for a feature is calculated from all the intra-CVs of all users. A low M-intra-CV for a feature means that the feature has low variability - in other words high stability - within all sessions for all users. Figure 5.4 shows the calculated M-intra-CVs for all the features (static and dynamic) arranged in a ranked order of stability or ease of repeatability for both the original and new signatures. A mean inter-CV (M-inter-CV) is also calculated for a feature from all inter-CVs across all users, where a low value of M-inter-CV shows that the feature is repeatable (or reproducible) for all users across all the capture sessions indicating a characteristic which does not fluctuate between the capture sessions.

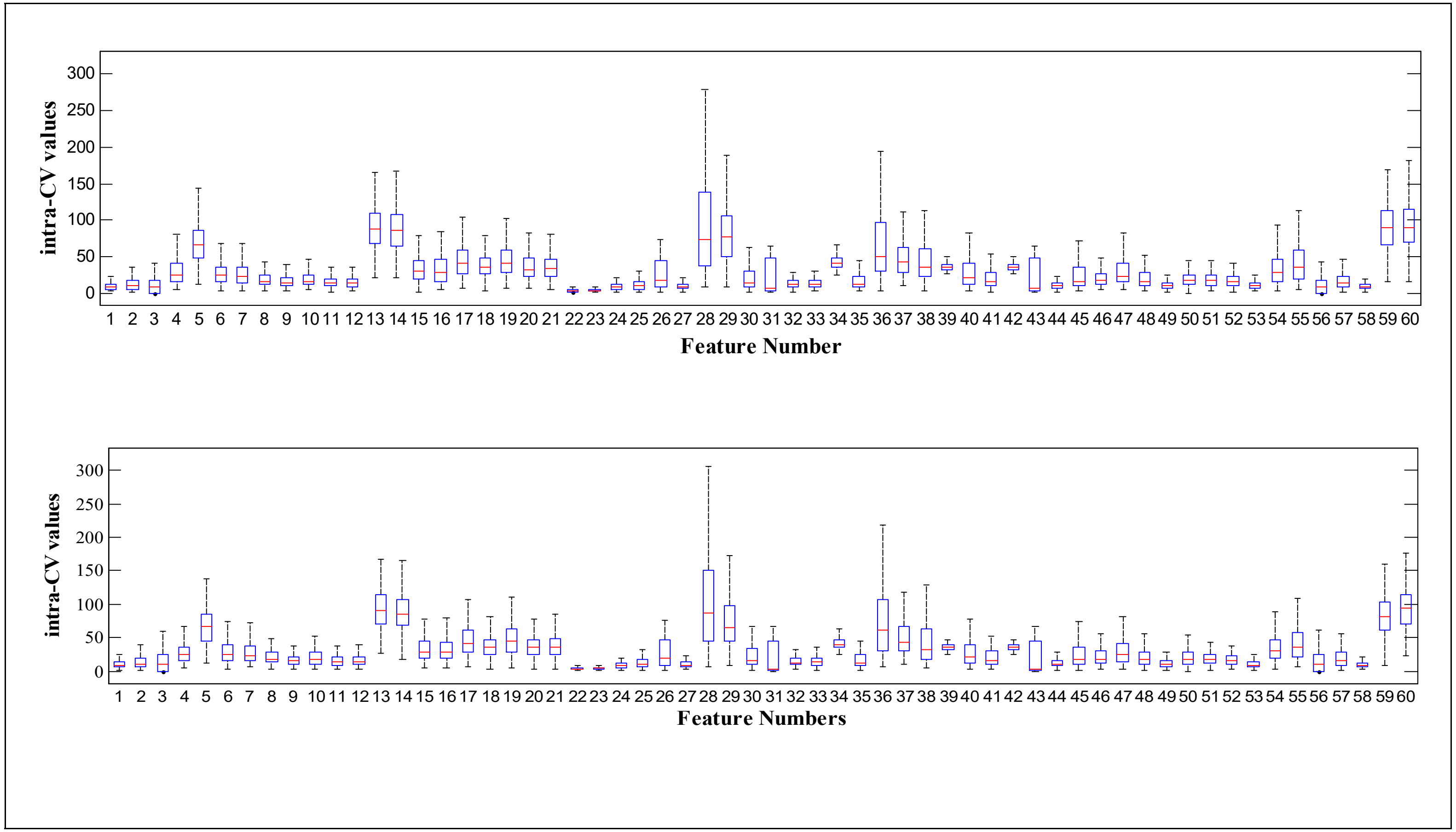


Figure 5.2. Intra session variation of individual feature in oringinal (top) and new (bottom) signatures

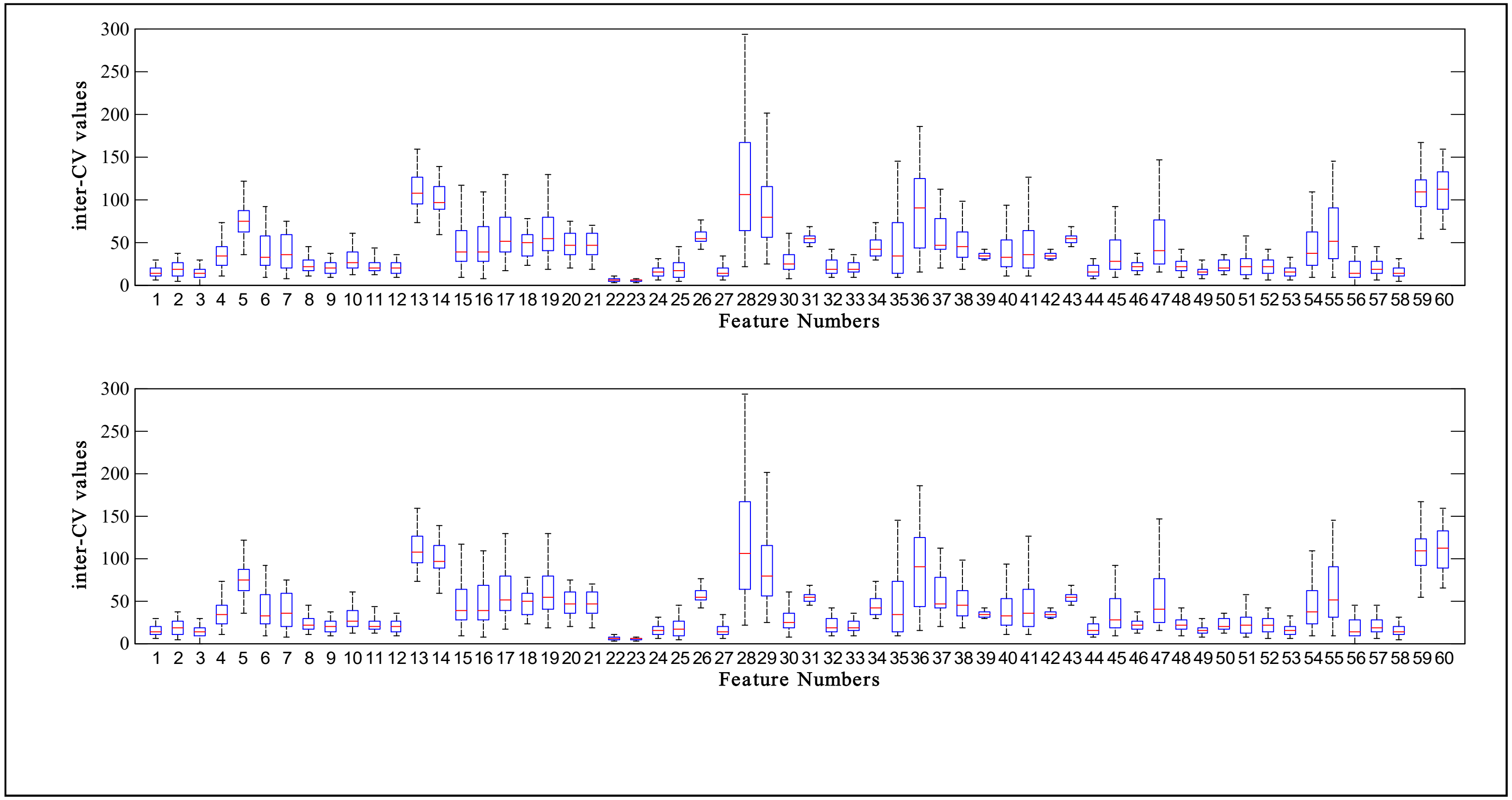


Figure 5.3. Inter session variation of individual feature in oringinal (top) and new (bottom) signatures

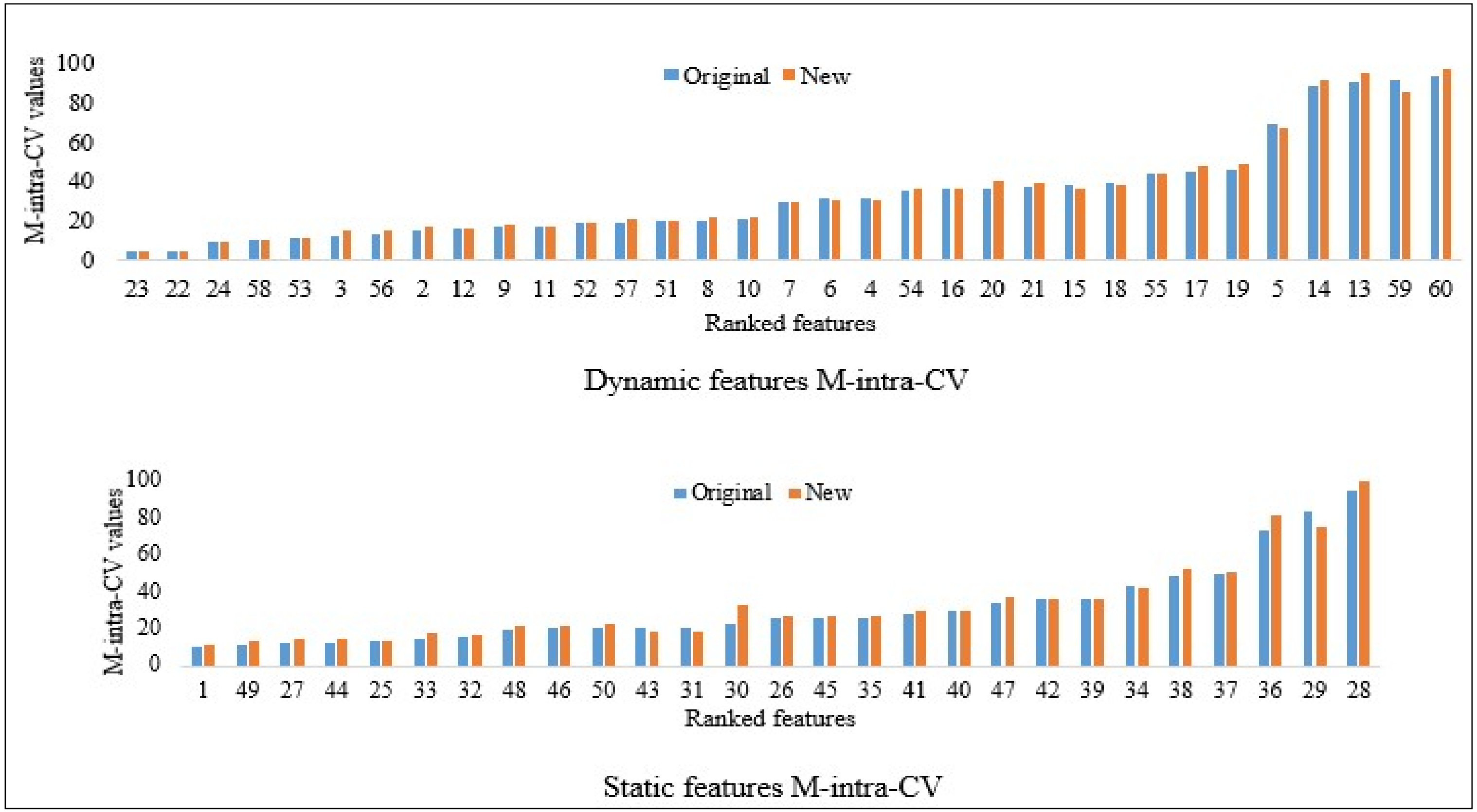


Figure 5.4. Mean intra-session variation of features across all users in a ranked order.

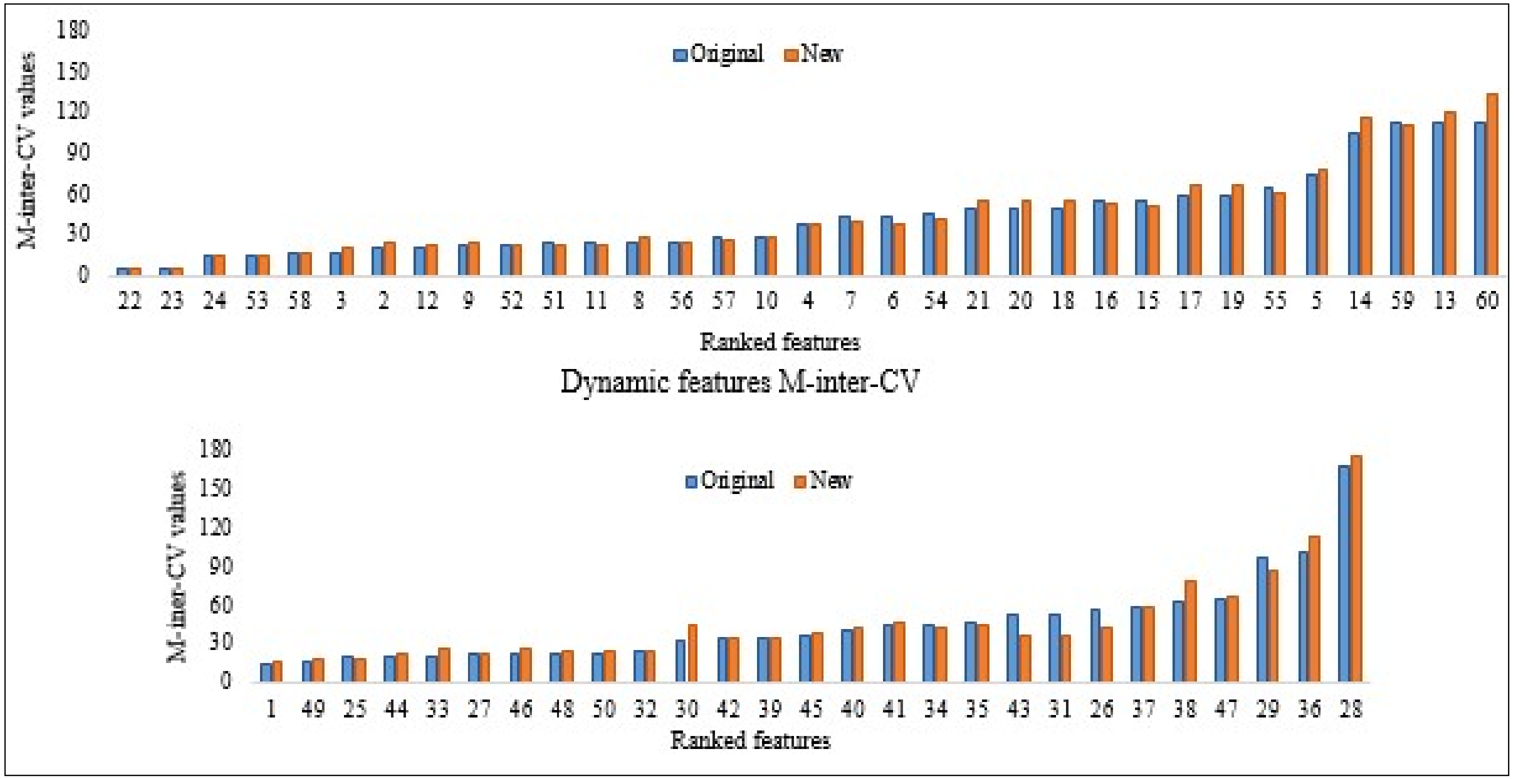


Figure 5.5. Mean inter-session variation of features across all users in a ranked order

Table 5.2. Top 20 most stable (least variable features)

Feature Number	Feature Name	Feature Type
22	Azimuth	Dynamic
23	Altitude	Dynamic
24	Pressure	Dynamic
53	Average resultant velocity	Dynamic
1	Total distance of pen travelled	Static
58	Average pen acceleration in resultant	Dynamic
49	height / area	Static
25	Number of points comprising the image	Static
3	Pen lift	Dynamic
44	Width of signature	Static
12	Maximum pen velocity in x - Minimum pen velocity in y	Dynamic
52	Total time of zero velocity / total time in Y direction	Dynamic
27	Standard deviation x coordinate values	Static
9	Maximum pen velocity in x - Minimum pen velocity in x	Dynamic
2	Total signature execution time	Dynamic
51	Total time of zero velocity/execution time in X direction	Dynamic
11	Maximum pen velocity in y - Minimum pen velocity in y	Dynamic
33	Average x coordinate value - minimum x coordinate value	Static
50	Number of vertical midpoint crossing the signature	Static
48	Signature width / area	Static

Figure 5.5 shows the calculated M-inter-CVs for all the features (static and dynamic) arranged in a ranked order of high to low repeatability (reproducibility) between sessions.

It can be noted that the highly stable or least variable features in the original signatures have similar mean values (both M-intra-CV and M-inter-CV) as the highly stable features in new signatures and a majority of the features' mean values are found to have a plus or minus value of 10 between the calculated mean values for original and new signatures indicating the stability or variability characteristics of the features remain almost consistent even when creating a new signature. It is also observed that the static features have higher variability compared to the dynamic features and the feature ranking is almost same for M-intra-CV and M-inter-CV in both original and new signatures. The top twenty least variable (or most stable) features are shown in Table 5.2 (other features' descriptions are given in Table 2.1, Table 2.2 and Table 2.3 in Chapter 2).

In order to investigate the effect of a feature's calculated value on its variability/stability/reproducibility (over a period), a Spearman rank correlation (as defined in Equation 2.2 in Chapter 2) is determined between the calculated mean feature values and the inter-CVs of each feature for every user for both original and new signatures. A significance level (p-value) is also calculated to determine the confidence in the relationship (as described in [287]) between feature values and their variability. The results of the correlation (correlation coefficient *rho* (*ρ)* and its significance level *p-value)* are shown in Table 5.3 and Table 5.4. (Table 5.5 shows the correlation results including all samples from both original and new signatures).

Based on the calculated correlation coefficient *rho(ρ)* and the significance level *p-value,* it can be observed in Tables 5.3 and 5.4 that several features are found to have significant correlation between the feature values and their variability, which indicates that the variability is influenced by the physical characteristics of the signature and/or temporal characteristics of the signing behaviour. Feature numbers 42 (Horizontal centralness), 43 (Vertical centralness), 31 (Average x-coordinate value), 39 (Average y-coordinate value) are found to have very strong negative correlations with their variability in both the original and new signatures being significant at $p<0.001$. This means that the variability of these features decreases as the feature values increase or, in other words, the higher the feature value the more reproducible the feature.

Table 5.3. Correlation between feature values and feature variability in original signatures (NS = Not significant)

Feature No	*rho(ρ)*	*p-value*		Feature No	*rho(ρ)*	*p-value*
1	-0.1626	NS		**31**	-0.7063	p<0.001
2	-0.2571	NS		**32**	-0.0474	NS
3	-0.3630	p<0.05		**33**	-0.4395	p<0.01
4	-0.5555	p<0.001		**34**	-0.5016	p<0.01
5	0.3478	p<0.05		**35**	0.1419	NS
6	-0.4972	p<0.01		**36**	-0.6915	p<0.001
7	-0.4597	p<0.01		**37**	-0.3188	p<0.05
8	-0.0868	NS		**38**	-0.6607	p<0.001
9	-0.3417	p<0.05		**39**	-0.9698	p<0.001
10	-0.2441	NS		**40**	-0.2870	NS
11	-0.2484	NS		**41**	-0.0391	NS
12	-0.2180	NS		**42**	-0.9698	p<0.001
13	0.2067	NS		**43**	-0.7063	p<0.001
14	0.7285	p<0.001		**44**	-0.2508	NS
15	-0.6271	p<0.001		**45**	-0.1233	NS
16	-0.5962	p<0.001		**46**	0.1838	NS
17	-0.2202	NS		**47**	0.1014	NS
18	-0.0921	NS		**48**	0.3799	p<0.05
19	0.2020	NS		**49**	0.2412	NS
20	0.1974	NS		**50**	-0.4800	p<0.01
21	-0.0326	NS		**51**	-0.4587	p<0.01
22	-0.5241	p<0.001		**52**	-0.3271	p<0.05
23	0.0818	NS		**53**	0.1921	NS
24	-0.5512	p<0.001		**54**	-0.3848	p<0.05
25	-0.4581	p<0.01		**55**	-0.4727	p<0.01
26	-0.2897	NS		**56**	-0.3190	p<0.05
27	-0.1057	NS		**57**	-0.3889	p<0.05
28	-0.6785	p<0.001		**58**	0.2271	NS
29	-0.5996	p<0.001		**59**	0.2409	NS
30	-0.4753	p<0.01		**60**	0.4030	p<0.05

Table 5.4. Correlation between feature values and feature variability in new signatures (NS = Not significant)

Feature No	*rho(ρ)*	*p-value*		Feature No	*rho(ρ)*	*p-value*
1	-0.0593	NS		**31**	-0.8004	p<0.001
2	0.2407	NS		**32**	-0.0717	NS
3	-0.1990	NS		**33**	-0.3221	p<0.05
4	-0.4802	p<0.01		**34**	-0.0729	NS
5	0.0945	NS		**35**	0.1769	NS
6	-0.0644	NS		**36**	-0.6842	p<0.001
7	0.0158	NS		**37**	-0.2283	NS
8	-0.0233	NS		**38**	-0.6221	p<0.001
9	-0.1036	NS		**39**	-0.9423	p<0.001
10	-0.0417	NS		**40**	-0.1719	NS
11	-0.2789	NS		**41**	0.1676	NS
12	-0.1034	NS		**42**	-0.9423	p<0.001
13	0.2231	NS		**43**	-0.8004	p<0.001
14	0.6055	p<0.001		**44**	-0.2492	NS
15	-0.2573	NS		**45**	0.0089	NS
16	-0.2362	NS		**46**	0.0652	NS
17	-0.3816	p<0.05		**47**	0.1767	NS
18	-0.0964	NS		**48**	0.3375	p<0.05
19	-0.0176	NS		**49**	0.3226	p<0.05
20	-0.1338	NS		**50**	-0.4186	p<0.01
21	-0.2702	NS		**51**	-0.2808	NS
22	-0.5273	p<0.001		**52**	-0.1834	NS
23	-0.0178	NS		**53**	0.0777	NS
24	-0.6466	p<0.001		**54**	-0.2565	NS
25	-0.0010	NS		**55**	-0.0245	NS
26	-0.3583	p<0.05		**56**	-0.2599	NS
27	-0.1429	NS		**57**	-0.4308	p<0.01
28	-0.8051	p<0.001		**58**	0.1745	NS
29	-0.4796	p<0.01		**59**	0.1983	NS
30	-0.4079	p<0.05		**60**	0.5617	p<0.001

Features 22 (Azimuth) and 24 (Pen pressure) are also found to have negative correlations, both being significant at $p<0.001$ in both original and new signatures - i.e. the higher the amount of pressure applied in signing, the more consistent is the user in applying that pressure. Signatures that cross the vertical centre point (Feature 50) more are found to be less variable than those that cross less frequently (significant at $p<0.01$ in the original and in the new $p<0.001$). Variability in the width of the signature does not seem to correlate its feature value significantly in original or new signatures. It is also observed that the more pen lifts within a signature, the easier it is to reproduce those lifts (significant at $p<0.5$ in original signatures but not in new signatures); the higher horizontal velocity applied while signing, the more consistent is the signer in applying those velocities (significant at $p<0.001$ in original signature and $p<0.01$ in new signatures).

A further study, to investigate if there is any effect of signature styles or external factors such as gender, age etc. on the features' variability, is reported in the next section.

5.3 Signature style and other factors

This section will present some analysis of the signature "style" adopted by signers when creating both the original and new signatures, exploring whether users typically tend to write the new signature keeping the same style as the original, or whether they adopt a different style; if there is any effect of signature style on original and new signatures based on individual feature values, its variability. Effects of other factors, such as age, gender etc., on both the original and new signatures are also explored later in this section. According to the studies reported in [177], [214], [282] signatures can be classified into three basic styles based on the legibility; i) *text-based(T)* - where all the allographs in the signature are clearly legible, *ii) mixed(M)* - where some of the allographs can be recognised but it is not possible to extract all of them completely, and *iii) stylised(S)* - where none of the allographs in the signature is legible. For this experimental study, all users' signatures, both original and newly established, are classified into these three style categories. The categorisation is carried out by using a small group of human observer to make a judgement. Five observers (university students and researchers) were asked to view a sample for each of all users' signatures and classify each sample as one of the three signature style categories (T, M, S). In this way, the style which was assessed as appropriate by a majority of the observers was finally chosen to be the designated style category for that signature.

Table 5.5. Percentage of users using each signature style

<table>
<tr><td></td><td></td><td>Original</td><td>New</td><td></td></tr>
<tr><td></td><td>Text-based(T)</td><td>23%</td><td>23%</td><td></td></tr>
<tr><td></td><td>Mixed(M)</td><td>33%</td><td>26%</td><td></td></tr>
<tr><td></td><td>Stylised(S)</td><td>44%</td><td>51%</td><td></td></tr>
<tr><td colspan="2" rowspan="2">Consistency of style</td><td colspan="3">Original</td></tr>
<tr><td>Text-based(T)</td><td>Mixed(M)</td><td>Stylised(S)</td></tr>
<tr><td rowspan="3">New</td><td>Text-based(T)</td><td>15%</td><td>5%</td><td>3%</td></tr>
<tr><td>Mixed(M)</td><td>3%</td><td>15%</td><td>8%</td></tr>
<tr><td>Stylised(S)</td><td>5%</td><td>13%</td><td>33%</td></tr>
<tr><td colspan="3">Total consistency</td><td colspan="2">63%</td></tr>
</table>

Table 5.5 shows the percentage of users using each style when signing their original and the new signatures. Overall, the stylised (S) signature style (original – 44%, new – 51%) is adopted more than the text-based (23% in both original and new) and the mixed (original - 33%, new - 26%) styles in both original and new signatures. 63% of the users used the same signature style as their original signature when creating the new signature and 36% of the users changed their signature style; where 3% and 5% of them changed style from text-based to mixed and stylised respectively. A change of style from mixed to text-based and stylised is observed

in 5% and 13% of the users; stylised to text-based and mixed is observed in 3% and 8% of the users respectively. This shows that the majority of the users tend to use the same signing style when creating a new signature.

5.3.1 Signature style influence

The effects of handwriting style (on feature values and feature variability) in original and new signatures are examined with two separate 3 × 2 (Style (T, M, S) × Type (original-O, new-N signatures)) analysis of variances (ANOVAs) for each feature and each feature's variability. Table 5.7 shows a summary of the ANOVA results for feature values (no significant difference is found in interactions between style and type of signature for any feature, and these are therefore omitted from the Table). Post hoc analyses are carried out with Tukey's HSD (honest significance difference) [290] to determine the style or type contributing to the difference if a significant difference is found.

It can be observed from the ANOVA result (in Table 5.6) for the comparison of feature 1 (total distance travelled by the pen) that no significant main effect is found for style, but significant difference ($p<0.05$) is found for type (original and new); where original signature traces found to have longer path lengths than the new signatures. For signature execution time (feature 2), the main effects for style and type are found to be significant (both *ps* <0.05). Post hoc analysis shows that text-based signatures take longer to execute than stylised signatures, where there is not much difference between mixed and stylised and mixed and text-based; also, original signatures take more time than the new signatures. A similar result is also found for pen lift (feature 3); users who use the text-based style lift the pen more than users who adopt the stylised style and the number of pen lifts is found to be greater in original signatures than new signatures. Likewise, total pen-up time (56) is also longer in text-based original signatures than stylised original and new signatures ($p<0.05$ for style and $p<0.01$ for type).

Table 5.6. ANOVA results for style (T, M, S) versus type (Original and new) for feature variabilities

F no	Style		Type		F no	Style		Type	
	$F_{(2,72)}$	p	$F_{(2,72)}$	p		$F_{(2,72)}$	p	$F_{(2,72)}$	p
1	0.52	NS	4.05	<0.05	**31**	0.24	NS	105.05	<0.001
2	3.68	<0.05	5.13	<0.05	**32**	0.37	NS	0.28	NS
3	7.27	<0.01	9.05	<0.01	**33**	1.79	NS	0.44	NS
4	0.28	NS	0.17	NS	**34**	5.85	<0.01	5.65	<0.05
5	0.29	NS	0.00	NS	**35**	9.20	<0.001	0.17	NS
6	4.37	<0.05	2.33	NS	**36**	3.49	<0.05	0.81	NS
7	4.46	<0.05	2.28	NS	**37**	0.20	NS	0.06	NS
8	4.67	<0.05	0.00	NS	**38**	1.80	NS	0.69	NS
9	6.57	<0.01	0.02	NS	**39**	1.00	NS	0.01	NS
10	4.06	<0.05	0.54	NS	**40**	2.02	NS	0.37	NS
11	6.57	<0.01	0.54	NS	**41**	4.68	<0.05	0.43	NS
12	7.90	<0.001	0.07	NS	**42**	1.00	NS	0.01	NS
13	0.73	NS	0.11	NS	**43**	0.24	NS	105.05	<0.001
14	0.60	NS	0.07	NS	**44**	0.90	NS	0.40	NS
15	3.91	<0.05	2.26	NS	**45**	4.64	<0.05	0.00	NS
16	4.00	<0.05	2.22	NS	**46**	3.83	<0.05	0.49	NS
17	0.16	NS	0.10	NS	**47**	0.54	NS	0.12	NS
18	0.23	NS	0.22	NS	**48**	6.29	<0.01	0.01	NS
19	0.02	NS	0.03	NS	**49**	2.38	NS	0.02	NS
20	0.13	NS	0.19	NS	**50**	4.04	<0.05	4.12	<0.05
21	0.15	NS	0.00	NS	**51**	6.54	<0.01	0.31	NS
22	4.19	<0.05	0.53	NS	**52**	4.67	<0.05	0.09	NS
23	4.79	<0.05	0.38	NS	**53**	10.64	<0.0001	0.01	NS
24	11.99	<0.0001	0.00	NS	**54**	5.00	<0.01	0.49	NS
25	5.92	<0.01	5.05	<0.05	**55**	3.84	<0.05	2.00	NS
26	6.15	<0.01	0.04	NS	**56**	3.47	<0.05	7.15	<0.01
27	1.36	NS	0.67	NS	**57**	1.22	NS	1.41	NS
28	0.35	NS	0.26	NS	**58**	2.56	NS	0.06	NS
29	0.67	NS	1.29	NS	**59**	0.93	NS	0.04	NS
30	1.46	NS	0.76	NS	**60**	0.54	NS	0.12	NS

No significant main effect is found either for style or type for average horizontal and vertical velocity (feature 4 and 5). But the main effect for style for average resultant velocity (feature 53) is significant at $p<0.0001$, and the results show that users sign at a lower velocity when using text-based style compared to when adopting stylised and mixed style. Number of zero velocities and accelerations in both horizontal and vertical direction (features 6,7,15,116) are higher in text-based original signatures than mixed and stylised original signatures (all significant at $p<0.05$). There is no significant difference found between original and new signatures for these features.

For azimuth and altitude (22 and 23), no significant main effect is found for type (original and new) but a significant (both *ps*<0.05) main effect is found for style. Mixed style signatures show higher azimuth compared to stylised with no significant difference between stylised and text-based. Text-based signatures reveal lower altitude compared to stylised with no significant difference between text-based and mixed or mixed and stylised. There is no significant difference in pen pressure observed between original and new signatures but a significant difference ($p<0.0001$) between stylised and text-based signatures is observed. Users apply higher pen pressure when writing stylised signatures than text-based or mixed.

The number of points comprising the signature image is significantly greater ($p<0.01$ for style and $p<0.05$ for type) in text-based original signatures than stylised original, mixed new and stylised new signatures with no significant difference between mixed and stylised. A similar difference is also observed for feature 34 (sum of y coordinate values significant at $p<0.01$ for style and $p<0.05$ for type) while the sum of x coordinates values (feature 26) is larger in text-based original signatures than stylised original signatures ($p< 0.05$ for style and NS for type) with not much difference between different styles in the new signatures. Mixed style original signatures cross the vertical midpoint more times than stylised original and stylised new signatures (significant at $p<0.05$ for style and $p<0.05$ for type).

Table 5.7. ANOVA results for style (T, M, S) versus type (Original and new) for feature variability

F No	Style		Type		F No	Style		Type	
	F(2,72)	p	F(2,72)	p		F(2,72)	p	F(2,72)	p
1	3.21	p<0.05	0.30	NS	**31**	2.52	NS	72.93	p<0.0001
2	2.72	NS	0.16	NS	**32**	6.00	p<0.01	0.17	NS
3	0.84	NS	0.98	NS	**33**	2.68	NS	1.23	NS
4	0.63	NS	0.13	NS	**34**	3.61	p<0.05	0.48	NS
5	0.21	NS	0.31	NS	**35**	0.81	NS	0.17	NS
6	4.33	p<0.05	0.51	NS	**36**	2.22	NS	0.69	NS
7	3.18	p<0.05	0.25	NS	**37**	0.48	NS	0.17	NS
8	0.07	NS	1.05	NS	**38**	0.94	NS	0.67	NS
9	0.67	NS	0.83	NS	**39**	0.47	NS	0.06	NS
10	0.06	NS	0.17	NS	**40**	1.74	NS	0.00	NS
11	0.69	NS	1.37	NS	**41**	0.83	NS	0.00	NS
12	0.85	NS	0.04	NS	**42**	0.47	NS	0.06	NS
13	0.78	NS	0.19	NS	**43**	2.52	NS	72.93	p<0.0001
14	0.70	NS	0.07	NS	**44**	6.11	p<0.01	0.48	NS
15	4.08	p<0.05	0.05	NS	**45**	1.30	NS	0.00	NS
16	2.97	NS	0.00	NS	**46**	0.32	NS	1.93	NS
17	5.04	p<0.01	1.45	NS	**47**	3.29	p<0.05	0.01	NS
18	2.17	NS	1.19	NS	**48**	1.49	NS	0.41	NS
19	3.88	p<0.05	1.26	NS	**49**	4.09	p<0.05	4.06	p<0.05
20	3.30	p<0.05	0.60	NS	**50**	1.44	NS	0.18	NS
21	2.48	NS	0.95	NS	**51**	7.00	p<0.01	0.03	NS
22	5.48	p<0.01	0.44	NS	**52**	4.51	p<0.05	0.00	NS
23	2.90	NS	0.07	NS	**53**	8.52	p<0.001	0.02	NS
24	3.45	p<0.05	0.04	NS	**54**	4.52	p<0.05	0.00	NS
25	4.81	p<0.05	0.41	NS	**55**	2.70	NS	0.02	NS
26	3.84	p<0.05	53.70	p<0.0001	**56**	1.67	NS	0.06	NS
27	6.40	p<0.01	0.10	NS	**57**	1.52	NS	0.19	NS
28	0.03	NS	0.01	NS	**58**	7.62	p<0.001	0.24	NS
29	0.18	NS	1.50	NS	**59**	0.03	NS	0.05	NS
30	0.61	NS	2.69	NS	**60**	0.43	NS	0.24	NS

Table 5.7 shows a summary of the ANOVA results for feature variability (no significant difference is found in interactions between style and type of signature for any feature, and these figures are hence omitted from the Table). As mentioned earlier, post hoc analyses are also carried out with Tukey's HSD to determine the style or type contributing to the difference if a significant difference is found.

It can be observed from the results (Table 5.7) that for many features the main effects for style and type (especially difference between original and new) are not significant or have little significance. For total distance of pen travelled, the main effect for style is marginally significant ($p<0.05$), where variability is higher in stylised original signatures than mixed style original signatures with no significant difference in variability between text-based new, mixed new, stylised new and text-based original signatures.

For azimuth (22), stylised original and new signatures vary more than text-based original and new signatures, with main effect for style significant at $p<0.01$. No significant difference in the variability is observed between original and new signatures. Pen pressure is found to be more variable in mixed style original than stylised original signatures ($p<0.05$), but the differences between the three styles in new signatures are not significant.

The number of points comprising the signature image varies more in stylised original than text-based original and mixed original signatures ($p<0.01$). The variability in stylised new is also higher than text-based new and mixed new but not significantly different. Sum of x coordinate values vary more in original signatures than new ($p<0.0001$ for type) where the variations in stylised and mixed original signatures are significantly different than the variations in stylised, mixed and text-based new signatures ($p<0.05$ for style). Width of the signature is observed to vary more in stylised signature ($p<0.05$), but there is no significant difference in variation observed between original and new signatures. Average resultant velocity and acceleration (53 and 58) in stylised signature vary more than the velocity in text-based or mixed style signatures (both ps <0.001 for style).

The number of points comprising the signature image varies more in stylised original than text-based original and mixed original signatures ($p<0.01$). The variability in stylised new is also higher than text-based new and mixed new but not significantly different. Sum of x coordinate values vary more in original signatures than new ($p<0.0001$ for type) where the variations in stylised and mixed original signatures are significantly different than the variations in stylised, mixed and text-based new signatures ($p<0.05$ for style). Width of the signature is observed to vary more in stylised signature ($p<0.05$), but there is no significant difference in variation observed between original and new signatures. Average resultant velocity and acceleration (53 and 58) in stylised signature vary more than the velocity in text-based or mixed style signatures (both ps <0.001 for style).

It is seen from the analysis in Section 5.2 that most of the static features change when creating new signatures and some dynamic features (such as pen pressure, velocity related features) do not show much difference between the original and new signatures. The work reported in this section suggests that adopting a different signing style from the original signature style when creating a new signature can make a difference in the features values of those features which are found to be very similar between original and new versions, when keeping the same signature style as the original. For example, if a user chooses to adopt a stylised signing style for the new signature in contrast to a text-based original signature, this would likely to increase the amount of pen pressure applied, average writing velocity, altitude and decrease the number of pen-lifts, pen-up time, points comprising the signature image etc. Although it is dangerous to draw an absolute conclusion from the experimental results presented in this section because of the small number of subjects in the database, nevertheless the results do suggest that adopting a different signature style for the new naturally revoked signature would have an impact on changing not only the static features but also some dynamic features.

5.3.2 Other factors

In this section the effects of age, gender, and handedness on original and new signatures are examined. Users of both original and new signature datasets are divided in to three age groups (age group 1: 16-25 years, age group 2: 25-40 years, age group 3: 40-60 years) as described in Section 3.4 in Chapter 3. Users are also divided into two gender groups (Gd 1: Female, Gd 2: Male) and two handedness groups (Hn 1: right handed and Hn 2: left handed) as described in Chapter 3. Percentages of the population for each group are also shown in Figure 3.15-3.17 in Chapter 3.

To examine possible effects of these characteristics two separate 3(age) x 2(type) x 2(Gd) x 2(Hn) analyses of variance (ANOVA) are performed for individual features (feature values and feature variability separately). Post hoc analyses are also performed using Tukey's HSD as described earlier. The significance of the effect of these factors (p values) from the ANOVA results is shown in Table 5.8.

It can be observed from Table 5.8 that the effects of age, gender or handedness are not found to be significant for either original or new signatures for feature 1 (total distance travelled by the pen). For signature execution time (feature no 2), age group 3 (40-60) take ($p<0.05$ for age and $p<0.01$ for type) a longer time to sign their original signatures than age group 1 and 2 signing both original and new signatures; also, when signing the new signatures age group 3 take a shorter time compared to the time taken signing the original signature. No significant difference is found in the new signature for any age group. Although the execution time may depend on the shape and size related features of the signature, this may also suggest that people can execute a new signature fluently or rapidly even if they take longer time to sign their original signature.

Table 5.8. ANOVA results (p-values) for age x type x gender x handedness for individual feature

F No	Age	Type	Gd	Hn	F No	Age	Type	Gd	Hn
1	NS	NS	NS	NS	**31**	NS	<0.0001	NS	NS
2	<0.05	<0.01	NS	NS	**32**	<0.01	NS	<0.05	NS
3	NS	<0.05	NS	NS	**33**	NS	NS	NS	NS
4	NS	NS	NS	NS	**34**	NS	<0.01	NS	NS
5	NS	NS	NS	NS	**35**	NS	NS	NS	NS
6	<0.001	<0.01	<0.05	NS	**36**	NS	NS	NS	NS
7	<0.001	<0.01	NS	NS	**37**	NS	NS	NS	NS
8	NS	NS	<0.05	NS	**38**	<0.05	NS	NS	<0.01
9	NS	NS	NS	NS	**39**	NS	NS	<0.05	NS
10	<0.05	NS	NS	NS	**40**	NS	NS	NS	<0.05
11	<0.05	NS	NS	NS	**41**	NS	NS	NS	NS
12	NS	NS	<0.05	NS	**42**	NS	NS	<0.05	NS
13	NS	NS	NS	<0.01	**43**	NS	<0.0001	NS	NS
14	NS	NS	NS	NS	**44**	NS	NS	NS	NS
15	<0.01	<0.01	<0.05	NS	**45**	NS	NS	NS	NS
16	<0.001	<0.01	NS	NS	**46**	<0.05	NS	NS	NS
17	NS	NS	<0.05	NS	**47**	NS	NS	NS	NS
18	NS	NS	NS	NS	**48**	NS	NS	<0.05	NS
19	NS	NS	NS	NS	**49**	NS	NS	NS	NS
20	NS	NS	NS	NS	**50**	NS	<0.05	NS	NS
21	NS	NS	<0.05	NS	**51**	<0.0001	NS	<0.05	NS
22	<0.01	NS	NS	NS	**52**	<0.0001	NS	NS	NS
23	NS	NS	NS	NS	**53**	<0.05	NS	NS	NS
24	NS	NS	NS	NS	**54**	<0.0001	NS	NS	NS
25	<0.05	<0.01	NS	NS	**55**	<0.001	<0.01	NS	NS
26	<0.05	NS	NS	NS	**56**	NS	<0.05	NS	NS
27	NS	NS	NS	NS	**57**	NS	NS	NS	NS
28	NS	NS	NS	NS	**58**	NS	NS	NS	NS
29	NS	NS	NS	<0.05	**59**	NS	NS	NS	NS
30	<0.05	NS	NS	NS	**60**	NS	NS	NS	NS

No significant difference is found in pen lifts for age, gender or handedness but pen lifts are greater in original than in new signatures. The number of zero velocities and accelerations in x and y directions are found to be greater in 40-60 years old male users' original signatures compared to other age and gender groups for both original and new signatures; even 40-60 years old male users' new signatures have less number of zero velocities and accelerations (not significantly different but suggesting users become confident signing the new signature irrespective of age) and in general female users have less number of zero velocities and accelerations than male users. Features 51, 52, 54 (ratio of total time of zero velocities to total duration in horizontal, vertical and resultant respectively) are found to have significant main effects for age. Post hoc analysis reveals that ratios are significantly higher in age group 3 (40-60 years) compared to age group 1 (16-25 years), and 2 (25-40 years), which suggests as people get older they tend to stop or pause more when signing. No effect of gender, type or handedness is found to be significant for these features. A significant main effect for handedness is found for average horizontal acceleration ($p<0.01$), where right handed users sign with higher horizontal acceleration than left handed users.

There is no significant main effect for age, type, gender or handedness in altitude and pen pressure. Azimuth tends to be higher in the 25-40 age group than in the 16-25 age group where no significant difference between 40-60 years' age group and the other two age groups (25-40 and 16-25). For number of points comprising the signature image, again the 40-60 years' users tend to have more points than the other two age groups. But the number of points are more in their original signatures than in their new signatures.

As mentioned earlier in this section a 3(age)×2(type)×2(Gd)×2(Hn) analysis of variance (ANOVA) is performed to examine the effect of age, gender, type and handedness on feature variability. The significance values of the effect of these factors (p values) from the ANOVA results are shown in Table 5.9.

Table 5.9. ANOVA results (p-values) age x type x gender x handedness for individual feature's variability

F No	Age	Type	Gd	Hn	F No	Age	Type	Gd	Hn
1	NS	NS	NS	NS	**31**	<0.05	<0.0001	NS	NS
2	NS	NS	NS	NS	**32**	NS	NS	NS	NS
3	NS	NS	NS	NS	**33**	NS	NS	NS	NS
4	NS	NS	NS	NS	**34**	NS	NS	NS	NS
5	NS	NS	NS	NS	**35**	NS	NS	NS	NS
6	NS	NS	NS	NS	**36**	NS	NS	NS	<0.001
7	NS	NS	NS	NS	**37**	NS	NS	<0.05	NS
8	NS	NS	NS	<0.05	**38**	NS	NS	NS	NS
9	NS	NS	NS	NS	**39**	NS	NS	NS	NS
10	NS	NS	<0.05	NS	**40**	NS	NS	NS	NS
11	NS	<0.05	<0.01	NS	**41**	<0.01	NS	<0.05	NS
12	<0.05	NS	NS	NS	**42**	NS	NS	NS	NS
13	NS	NS	NS	<0.0001	**43**	<0.05	<0.0001	NS	NS
14	NS	NS	<0.05	NS	**44**	NS	NS	NS	NS
15	NS	NS	NS	NS	**45**	NS	NS	NS	NS
16	NS	NS	NS	NS	**46**	<0.01	NS	NS	NS
17	NS	NS	NS	NS	**47**	NS	NS	NS	NS
18	NS	NS	NS	NS	**48**	<0.01	NS	NS	NS
19	NS	NS	NS	NS	**49**	NS	NS	NS	NS
20	NS	NS	NS	NS	**50**	<0.05	NS	NS	NS
21	NS	NS	NS	NS	**51**	NS	NS	NS	NS
22	NS	NS	NS	NS	**52**	NS	NS	NS	NS
23	NS	NS	NS	NS	**53**	NS	NS	NS	NS
24	NS	NS	<0.01	NS	**54**	NS	NS	NS	NS
25	NS	NS	NS	NS	**55**	NS	NS	NS	NS
26	NS	<0.0001	NS	NS	**56**	NS	NS	NS	NS
27	NS	NS	NS	NS	**57**	NS	NS	NS	NS
28	<0.0001	NS	<0.01	NS	**58**	NS	NS	NS	NS
29	NS	NS	NS	<0.05	**59**	NS	NS	NS	NS
30	NS	NS	NS	NS	**60**	NS	NS	<0.001	NS

It can be observed from the Table that not many features have a significant effect of age, type, gender, handedness on their variability. The variation in horizontal acceleration is higher in original signatures where users sign with their left hands (significant at $p<0.0001$). Vertical acceleration varies more in signatures written by male users compared to female users ($p<0.05$). But for pen pressure it is the opposite, the variation in pen pressure is higher in female than male signers: i.e. male users are better in reproducing the same pressure. Variation in average x coordinate value is significantly higher in original signatures written by users from age group 1 and age group 2 compared to new signatures written by users from same age groups. Signature width by height ratio tends to be more consistent in signatures written by 16-25 years old users than 40-60 years old users (significant at $p<0.01$). Also, the number of vertical midpoint crossings (feature 50) varies more in 40-60 years old users than 16-25 years old users ($p<0.05$). Both effects indicate that younger users reproduce size and shape related features better.

In this section the influence of age, gender and handedness on individual feature of original and new signatures has been analysed. It is found from the experimental results that female writers generally tend to stop or pause less than the male writers when signing. The older male users (40-60), especially, tend to stop more and for longer compared to younger male users (16-25 and 25-40) but interestingly the number of pen stops while signing becomes less when creating the new signatures compared to the original signatures for older users. Although this could be due to the size, shape or the style of the signature (as discussed in Section 5.3.1), this may suggest also that people are able to create a new signature fluently irrespective of their age, maybe by adopting a different signature style. Right-handed users are found to be able to sign faster and reproduce the acceleration better than the left-handed users. Younger users (16-25) are found to be better in reproducing size and shape related features than older users. Although it is hard to draw an absolute conclusion from the results because of the small database used for the experiments, the results showed that a majority of the features do not show much difference between original and new signatures as a function of age, gender and handedness in terms of their feature values and reproducibility, which indicates that the idea of natural revocability (adopting a new signature in the event of compromise) can be feasible for any age, gender and handedness as the effect of these factors are not very different from the original signature. Moreover, for some features the effect of ageing can be overcome in the naturally revoked signature by adopting a different signature style.

5.4 Conclusion

In this chapter, a feature based analysis of the natural revocability phenomenon has been presented by investigating the relationship between features, signature style and their effect in original signatures and new signature. A brief review of feature based studies in handwritten signatures in relation to feature variability, signature style and other demographic factors has initially been provided. Correlations and differences between original and new signatures have been explored by experimental analysis and discussed including feature values, feature variability and the relationship between feature values and variability. Many features have been found to be strongly correlated between original and new signatures. Both intra-session and inter-session variability has been analysed for each feature and the ranking of feature variability has been found almost identical in both original and new signatures, indicating that these characteristics are general handwritten signature characteristics. An analysis of different signature styles has been presented investigating consistency of signature style when changing to a new signature, effects of style on signature features between original and new signatures. Finally, the effects of demographic factors such as age, gender, and handedness on original and new signatures have been explored and discussed.

In summary, the study presented in this chapter provides a valuable insight into not only the feature based characteristics in naturally revoked signatures but also in general handwritten signatures.

The next chapter will investigate an objective measure of a new feature ('Hesitation') which may be useful in describing the stability in natural revocability as well as exploring its possibility to be an effective discriminating feature in automatic handwritten signature verification more generally.

Chapter 6:

Developing features to improve handwritten signature biometrics

---

This chapter will explore the development of a type of feature for signature processing which appears to be particular relevance to the study reported here. The feature relates specifically to the concept of hesitancy (or its converse, fluency) of and we will investigate its impact on signature development in the context of natural revocability and signature verification more generally, using an objective measure of the power of the feature. Section 6.1 will introduce the general idea of hesitation and its measurement and provide a brief review which covers all the relevant studies and background information about hesitation in different areas. Section 6.2 will introduce and describe an algorithm to measure the feature value quantitatively. Following the objective definition of hesitation some experimental analysis will be reported in section 6.3. Finally, section 6.4 will conclude the chapter.

6.1 Introduction

As discussed in Chapter 1, the handwritten signature has long been established as a common means for providing proof of identity, especially in the context of certification of documents, financial transactions, and so on, while, in many practical situations, everyday signature checking encountered in public services or sales environments involves non-experts with no formal training. In more formal situations, for example in legal scenarios of in criminal investigations, the determination of the authenticity of a handwritten signature is still most commonly carried out by human inspectors, usually professional forensic document examiners (FDE). Automatic signature verification systems can provide more robust solutions in the former situation, but can also support and enhance performance in the latter, with both online and offline processing scenarios encountered [17], [119], [198], [206]. Most of the work carried out by FDEs regarding signatures is focused on offline signatures (where the signature is in the form of an ink trace on a substrate, normally paper) and visually detectable features in handwritten signatures form the basis of evidence supporting whether a questioned signature is genuine or forged [214], [291]. But in this way, significant dynamic information (such as pen velocity, stroke duration, etc.) is lost to the examiner, although estimates of these dynamics may sometimes be inferred [292], [293] from static traces. It has also been reported in published studies [93], [294] that forgers are more successful in copying spatial aspects of handwriting than the kinematic aspects, as forging another person's handwriting or signature is an untrained motor task that needs feedback in order to simulate general shape characteristics in the best possible way. As a result, the writer often shows hesitation, unnatural pen lifts, patching, tremor etc. in writing or signing, revealing the true nature of forgery [197] .

Normally, the signature is written quickly in a fast and fluent motion, as this is probably the most practised handwriting movement developed in one's lifetime unless an individual is suffering from some pathology or other physical or mental condition that influences their handwriting performance [295]. These fast fluid movements contribute greatly to the flow of the signature and as they are frequently rehearsed, and become automatic to the writer [85], [90]. The more this is skilled and automatic, the less variability there is in the temporal domain (performance time), the spatial domain (length, height and width) and in characteristics relating to pressure (force applied on an object or towards a surface) measures, and the more consistency and fluency is evident [99]. When a forger imitates a signature these dynamic movements in signing are generally not imitated accurately, resulting in line breaks, hesitations, slow starts and endings and so on [296]. The literature shows that poor line quality characteristics have been classically associated with simulated movement and the ASTM Standard Guide for Examination of Handwritten Items [297] specifically lists certain characteristics such as "lifts, stops, hesitations" [294], as those of which FDEs should take special note in assessing the authenticity of a signature.

Some aspects of handwriting/pen movement have been studied based on observations of pen lifts, pen stops, velocity, pen force etc. in order to analyse these characteristics differentially in relation to genuine and forged or mimicked signatures [298] and based on these handwriting movements (fluency , level of tremor etc.) dynamic signature analysis tools [199] have been introduced to provide statistical support for forensic analysis; but little work has been reported on objectively measuring 'hesitation' specifically, even though this is one of the common signature characteristics in identifying forgeries in forensic practice and, by implication, is therefore likely to be of significance in developing a new signature in the context of natural revocability This chapter will investigate the possibility of defining some objective measures of hesitation and their influence on handwritten signature analysis (specifically on signature verification systems) by means of some experimental analysis. A brief survey of definitions and measures of fluency and hesitation as reported in the literature is provided next. As we will also show, relevant work in child development studies also supports the idea that this type of feature is likely to be significant from the viewpoint of our concept of natural revocability.

6.1.1 General notions of fluency and hesitation

According to the study reported in [299], the general meaning of fluency is defined as completing a task effortlessly, where a person writes the signature “automatically, fluidly, rapidly, quickly, and accurately”. In other literature, fluency is described by using similar terms such as effortless, without hesitation, fearless in making mistakes, etc. [299]–[301]. In terms of writing, it is the writer’s ability to write without excessive hesitations, blocks, and interruptions [301]. In the area of handwriting development, fluency is quantified according to the total number of letters, words or sentences written within a given time period (letters or words per minute). Studies in children’s handwriting development show [302] that handwriting fluency measured in this way increases with age progression. As fluency in writing is the end product of all the writing processes [303], an increased fluency in writing indicates more efficient writing processes and this fluency and ‘automisation’ result in reduced demands on working memory (i.e. decreased cognitive effort). This means fewer pauses (less hesitation), less variation in letter height and width, more spatial accuracy, and better control of pen pressure levels [304], [305].

It is noticeable from the studies reported in the literature that hesitation, on the other hand, is a form of disfluency or lack of fluency. Generally, it is a descriptor of human activity such that, whenever a person is not sure about a task, re-evaluation of the situation is performed. Psychologists describe this as the time elapsing between the external or internal stimulation of an organism and his, her, or its internal or external response [306], and can be expressed through facial expression, head movement, special verbal markers etc. [307][. From a psycholinguistic perspective [308] “hesitation disfluencies are found to occur more often before longer utterances and when the topic is unfamiliar.” and it is quantified as the number of pauses greater than 0.5 seconds when speaking [309].

Similarly, forensic researchers have defined hesitation as a pause or stop in the writing motion in which the writing instrument remains in contact with the writing surface [310] and this can be observed in the process of creating a forgery, as the forger may pause to consult the genuine signature and then continue duplicating it and sometimes it may leave an ink mark on the page (blobs) [174] .

The hesitation phenomenon has been extensively invoked in neuroscience to explain neurological disorders for patients walking, talking and writing [259]. It has also been used in testing visuo-motor dysfunction for a patient's rehabilitation process [311], [312], for handwriting analysis of patients with Parkinson's disease [97] and is also mentioned in [205] as a sign of constrained signature and can be useful for detecting forgery. The study reported in [311] was among the first where hesitation has been defined in a quantitative approach in drawing shapes.

It is clear from the review and discussion above that the hesitation phenomenon (i.e. the inverse of fluency) plays an important role in children's handwriting development, psychology, speech analysis, neuroscience, forensic document examination, and biometrics. Due to its importance, it is thus useful to seek objective and robust hesitation measures which can be used in providing automated processing tools. It is noted earlier that this is an area which has not been extensively studied formally, and so in the next sections of this chapter some objective yet simple ways of quantifying the notion of hesitation in handwriting are introduced, and the practical implications are investigated in relation to handwriting analysis.

6.2 Towards an objective measure of hesitation

Though the notion of *'hesitation'* is intuitive and natural to human analysts, this concept has not generally been objectively defined in a way which can unambiguously be embedded in algorithmic form for automated analysis. This section, therefore, explores some objective ways of quantifying the notion of hesitation in handwriting with a view to defining the concept in an algorithmic way. As noted in Section 6.1 the study reported in [311] is among the first where hesitation is defined objectively as the number of times the pen velocity falls beneath a threshold value during a single stroke. The work reported in [97] followed the same definition when examining handwriting analysis in patients with Parkinson's disease. There are clearly many ways in which hesitation can be algorithmically defined. Three very simple hesitation measures are introduced here as a starting point, and these are used to explore the role of this feature in handwritten signature analysis tasks. These three hesitation measures are based on common intuitive definitions of hesitation as discussed in Section 6.1.1; when a person is not confident or is hesitant while writing, the pen moves a relatively short distance and/or the pen becomes pretty much stationary. Consequently, in order to capture these intuitive notions, one of the defined measures is based on the amount of time during which the pen is approximately stationary, a second is based on the (short) distance travelled by the pen while moving very slowly (almost stationary) and the third measure is based on the time when the pen moves a very short distance. These measures are obviously likely to show a degree of correlation, but it is worth exploring these three algorithmically implemented intuitive hesitation measures as a starting point. The three simple measures considered here are defined as follows:

a. **H1**: This is simply a measure of the proportion of time the pen is stationary or near to stationary for the duration of a specified writing segment. Specifically, in this study, it is the ratio of the total time during which the pen is effectively at rest (not moving) or near to rest to the total time to execute the signature. This definition is close to the original definition found in [97], [311] and is specified in (6.1) below for the purpose of the present study:

$$Hesitation[H1] = \frac{t_v}{T} \quad \textbf{(6.1)}$$

Where t_v = total time when velocity < 2mm/sec [97], [311] and

T = total time to execute the signature.

b. **H2**: In this measure hesitation is defined as the proportion of the distance the pen travelled when the pen is moving below a very low velocity (2mm/sec) to the total distance travelled to execute the task and is defined in Equation 6.2.

$$Hesitation[H2] = \frac{d_v}{D} \quad \textbf{(6.2)}$$

Where d_v = Total distance travelled when the pen velocity<2mm/sec and

D = Total distance travelled to execute the task (signature).

c. **H3:** Here hesitation is defined as the proportion of time taken for the pen to travel a very short distance from one sample point to the next sample point (0.05mm-empirically chosen) or below that distance to the total time to execute the task. This measure is defined in Equation 6.3.

$$Hesitation[H3] = \frac{t_d}{T} \quad \textbf{(6.3)}$$

Where t_d = Total time when distance travelled <0.05mm and

T = Total time to execute the task

Note that these measures each require the specification of a particular threshold value. For hesitation measure H1 and H2 threshold value 2mm/sec was chosen on the basis of definition of very low pen velocity (or when the pen is almost stationery) reported in [311]. For the hesitation measure H3 when the pen travels a very short distance, we have experimented with changing threshold values and 0.05mm was determined experimentally such that a balance is achieved between the performance and the measured hesitation value.

The next section of this chapter will explore the influence of these three algorithmically extractable hesitation features on signature development and signature verification.

6.3 Experimental investigation of the objectively defined hesitation parameter

In this section, some experimental analysis is reported to investigate the influence of the objectively defined hesitation measures on handwritten signature analysis - particularly on handwritten signature development (when creating a new signature in the context of the natural revocability concept) and signature verification (verification of genuine signatures and rejection of forgeries). Signature samples captured in two databases – RevKent and BioKent (described in Chapters 3 and 2 respectively) are utilised for the experimental analysis. Details of the data collection protocols, procedures and storage of the RevKent database have been described in Chapter 3 and the BioKent database (where handwritten signature samples were collected as part of a multimodal database) have been described in Section 2.6.5 in Chapter 2. Both original and new signature samples in RevKent and genuine and forgery signature samples in BioKent) are utilised for the experimental analysis.

The features designated above as H1, H2, H3 are extracted from all samples in each of the two databases as defined in Equation 6.1, 6.2 and 6.3 and labelled as feature number 61(H1), 62(H2) and 63(H3) respectively. All of the hesitation features defined can take on a value from 0 to 1 where a value closer to zero signifies less hesitation, and a value closer to 1 a higher degree of hesitation. Thus, the greater the hesitation in signing, the greater the value of H returned.

6.3.1 Hesitation – in the context of natural revocability

As noted earlier in [85], [90], it is known that the individual handwritten signature becomes automatic to the writer because of rehearsing this regularly and frequently over a long period of time. Therefore, the user is expected to have better fluency, and thus lower hesitation, when signing the original signature which s/he has been signing since s/he started signing (learned and practised over a long time), but there is a question about what happens when an individual creates a new signature: does hesitation change with time (increase or decrease or is it not affected at all)? Another relevant question is: do all the different hesitation measures have the same effect or are some better indicators than others of the occurrence of hesitation? These basic research questions are investigated in this Section, based on experimentation utilising the RevKent database (as described in Chapter 2).

As described in Section 6.2 the three different hesitation features (H1, H2, and H3) are extracted from all samples for both original and new signatures contained in the RevKent database. In order to show the evolution of the hesitancies in signatures with time, statistical analysis is carried out across the sessions and these variations are measured in terms of their medians and 25th and 75th percentiles, where an increase or decrease in their medians will indicate increase or decrease in the hesitancies with time. First, the average of each hesitation feature is calculated for a given user's signature samples belonging to one session. As described in Chapter 2, majority of users completed at least four sessions, experimentation is initially carried out for these four capture sessions.

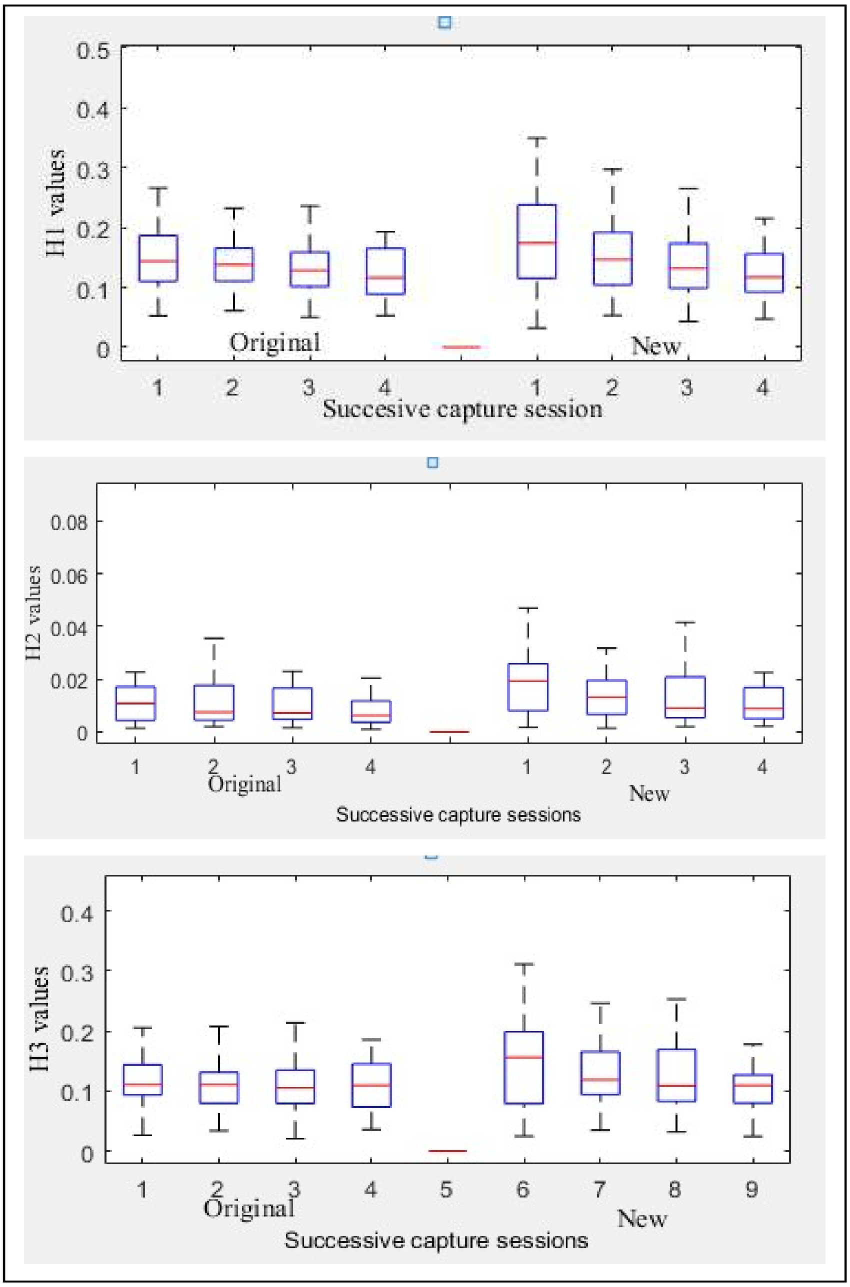


Figure 6.1. Boxplots of different hesitation values across 4 sessions in original and new signatures (at the top H1, middle H2, and bottom H3)

In Fig. 6.1, boxplots of the average hesitation values in successive sessions are depicted. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points observed.

It can be observed that all the hesitation measures H1, H2 and H3 show a downward trend, as we should perhaps expect, showing that signers (or users) become more fluent (or less hesitant) as time progresses. Though this downward trend is visible in both original and new signatures, this is more evident in the new signature.

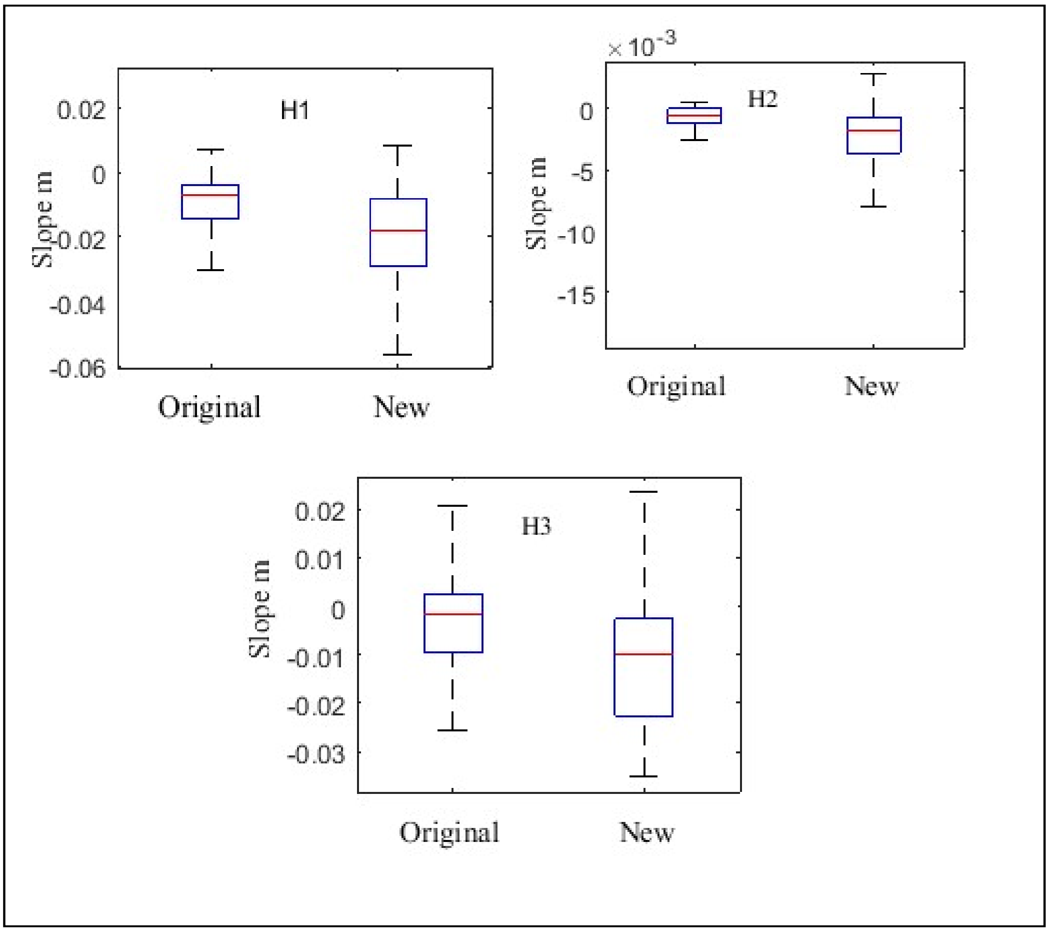


Figure 6.2. Variation in hesitation with time (measured by slope m) showing H1 at top left, H2 at top right and H3 at bottom

To obtain a clearer picture of these patterns, a linear regression model is developed for the average hesitation values for each user for each hesitation measure and the slope (*m*) of the regression line is measured for both the original and the new signature for each user as defined in (4.3) in Chapter 4. Each value of *m* obtained by the evaluation is either positive (above zero) or negative (below zero), and a positive value of *m* represents an increase in hesitation values with time while a negative value of m represents a decrease in hesitation values with time. The further is the value from zero, the steeper the regression (in this case the hesitation values change rapidly with time) either in a positive or a negative direction. The measured slope (*m*) values for each hesitation measure are shown in Figure 6.2 using box plots.

It can be observed from Figure 6.2 that the hesitation values in the new signatures decrease rapidly with time compared to the original signatures and the slight downward trend in the original signatures is most likely due to the effect of unfamiliarity of the signing system. But it is quite clear from Figure 6.1 that the hesitation values in the first session are higher in the new signatures compared to the same in the original signatures (as, obviously, the users are familiar with their original signature) and, more importantly, these reduce with time, providing evidence that the signers become less hesitant and more confident in signing the new signatures. In other words, the new signatures developed by the subjects do appear to stabilise as time progress.

This analysis is based on just four successive capture sessions. In order to explore this further, and as suggested by [313] that effective handwriting interventions require more practice, signatures were collected for 10 sessions from 9 individuals (those who were still available and willing to take part), the sessions taking place once a week. Figure 6.3 shows the variation of hesitation values across 10 sessions for those 9 individuals, again using box plots. In fact, the downward trend observed

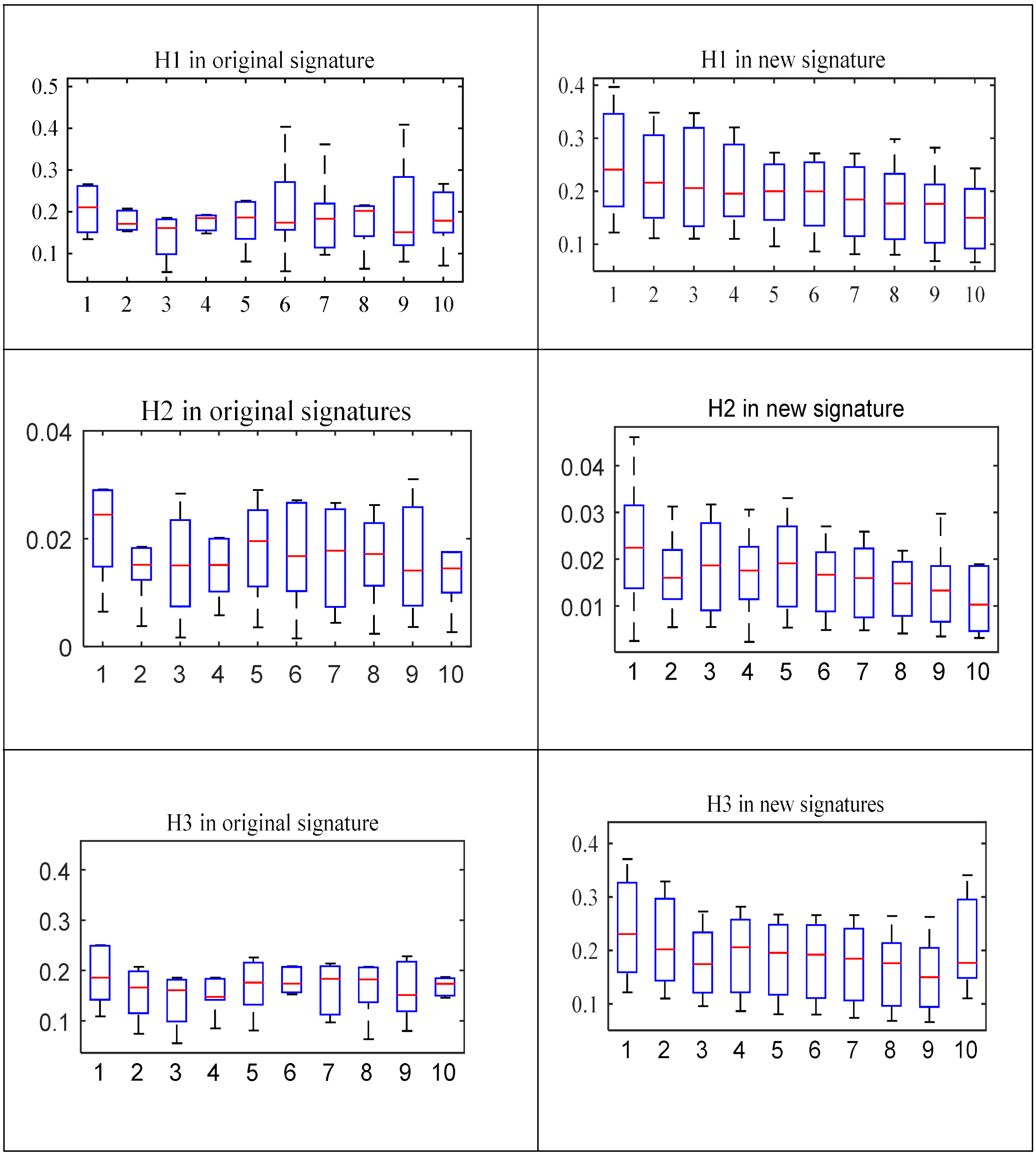


Figure 6.3. Boxplots of different hesitation (H1, H2, and H3) values across 10 sessions in original and new signatures

previously is more evident in here than when signatures were collected for just 4 sessions, which makes it clear that with time hesitation reduces, and that the new signatures achieve stability even over a relatively modest timescale. However, as mentioned earlier, even if it is considered that 10 sessions may not be enough to attain ultimate stability, these results are an indication that with time hesitation reduces. Statistically, it is accepted that 9 subjects for 10 sessions does not represent a large enough sample size to establish unequivocally that hesitation reduces with time (as its inverse, fluency, improves with time), but this preliminary indication supports the findings of the experimental analyses carried out in Chapter 4 that can lead to important research in handwritten signature development and further development in handwritten signature verification, and is a very important finding in the context of the concept of natural revocability.

6.3.2 Hesitation: relevance to analysis of forgeries

As mentioned previously, in assessing the authenticity of a signature, a human FDE compares and assesses features (such as stroke length and slant, letter formation, connecting strokes, pen lifts, line quality, pen pressure, base alignment, *hesitation*, patching etc.; patching is a flaw in the writing line in the form of a correction attempted by the forger to fix an obvious defect) between the questioned and known signatures and then makes a subjective judgment as to whether the signature is genuine or not. In this section, some experimental analysis is performed to compare the behaviour of different hesitation features in original and forged signatures, utilising the previously described BioKent database. Signature samples were collected from 79 subjects, where each of the subjects provided their signature samples in two sessions. Each subject donated 30 genuine (15 in each session) and 20 skilled forgery samples (10 in each session).

Reported human studies have shown that the value of hesitation is expected to be lower in genuine signatures because of the signer's obvious familiarity with the process [314]. Figure 6.4 shows a comparison between H-measures in the genuine and forged samples using box plots, and it is clear from these results that a majority of the forgery signature samples show much higher hesitation (in all three measures H1, H2 and H3) than the genuine signature samples, suggesting that when a person imitates someone else's signature he/she tends to be more hesitant in signing than the genuine signer. Also, Figure 6.5 shows a comparison of hesitation values in original and forged signatures for individuals, where it is evident that for H1, in 96% of users a forged signature has much higher hesitation than that of genuine signature; for H2, in 97% of users the forged signature has much higher hesitation than that of genuine signature; and for H3, again in 97% of users the forged signature shows much higher hesitation than that of the genuine signature. This in turn suggests that measuring hesitancy might be a very useful forensic tool to aid the automated discrimination between genuine and forged signature samples.

Another statistical analysis is carried out to observe the difference in hesitation measures between genuine and forged signatures using frequency histograms. Signature samples from the BioKent database are used for this investigation. Figure 6.6 shows the frequency histograms for each objectively defined hesitation features (H1-61, H2 – 62 and H3 – 63). The frequency histogram shows that all the hesitation features show lower values in the genuine signatures compared to the forgeries (as also shown in Figure 6.4 and 6.5). For the H1 feature the frequency histogram indicates (Figure 6.6) that a questioned signature is most likely to be a forged specimen (with 99.99% confidence, $p<0.0001$) if the value of the measured hesitation (H1) is more than 0.48. If the value of the hesitation(H2) in a questioned signature is more than 0.13, is most likely to be a forged counterpart (with 98% confidence, $p<0.02$). Figure 6.6 also indicates that a questioned signature is most likely to be a forgery if the value of H3 is more than 0.3 (with 99% confidence, $p<0.001$). This information could be very useful for automating forensic examinations or in providing automated tools to assist the human forensic examiner in the process of analysing genuine and forgery signatures captured online.

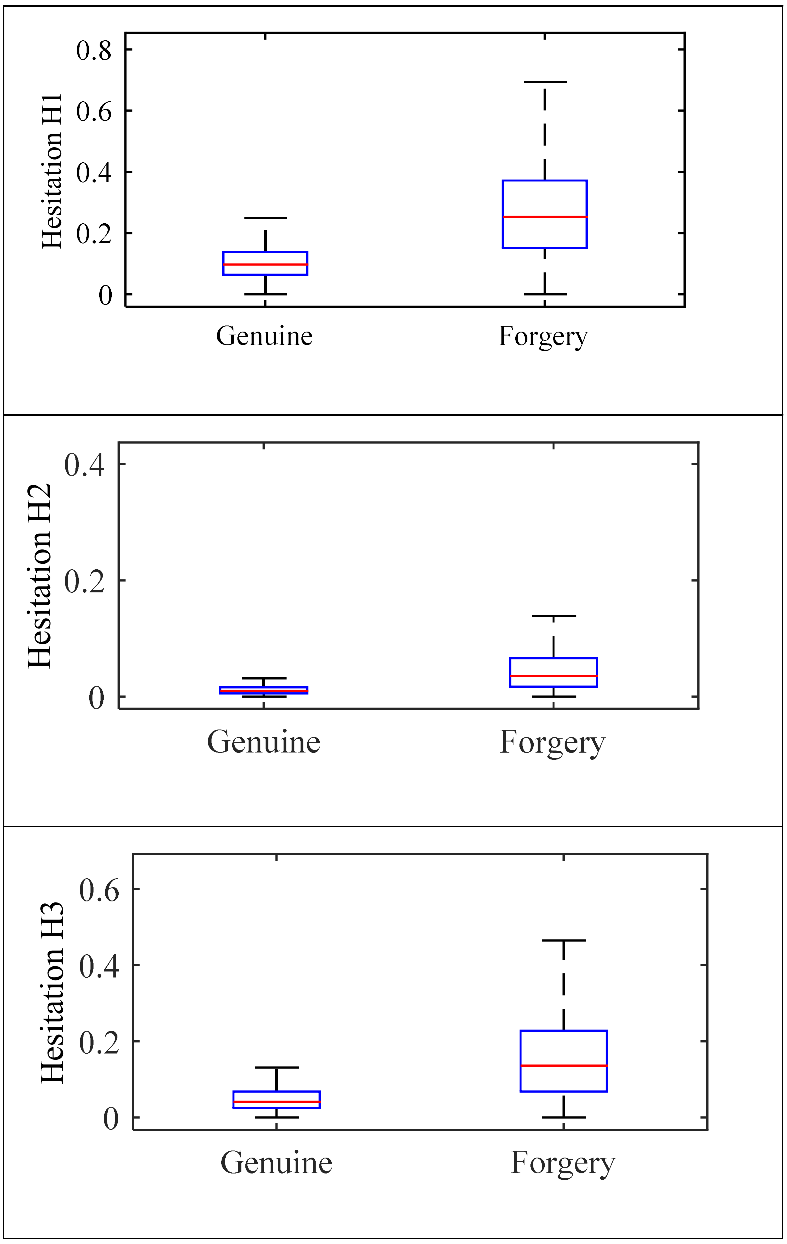


Figure 6.4 lots of different hesitation (H1, H2, and H3) values in original and forged signatures

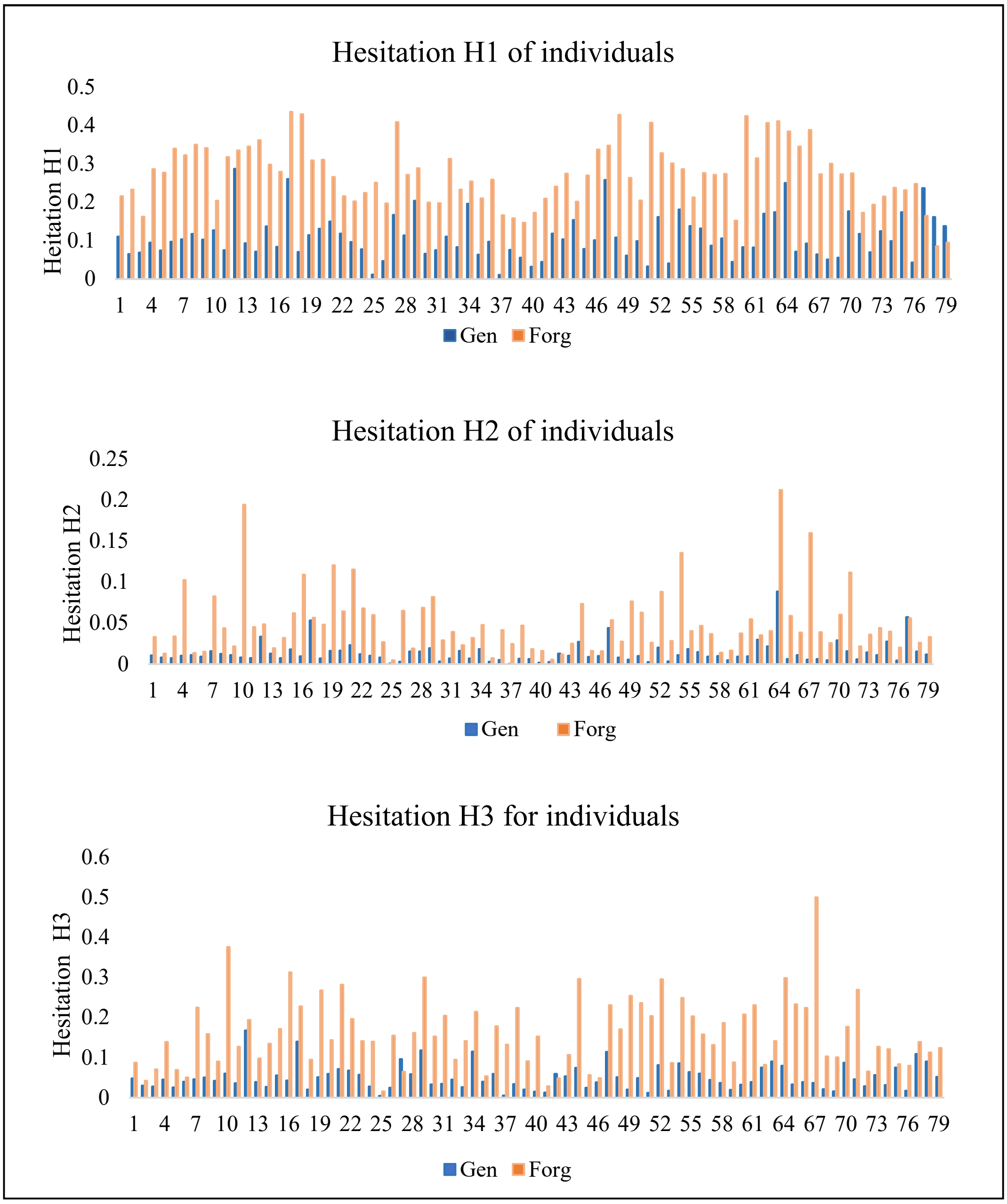


Figure 6.5. Different hesitation (H1, H2, and H3) values for individuals in genuine and forgery signatures

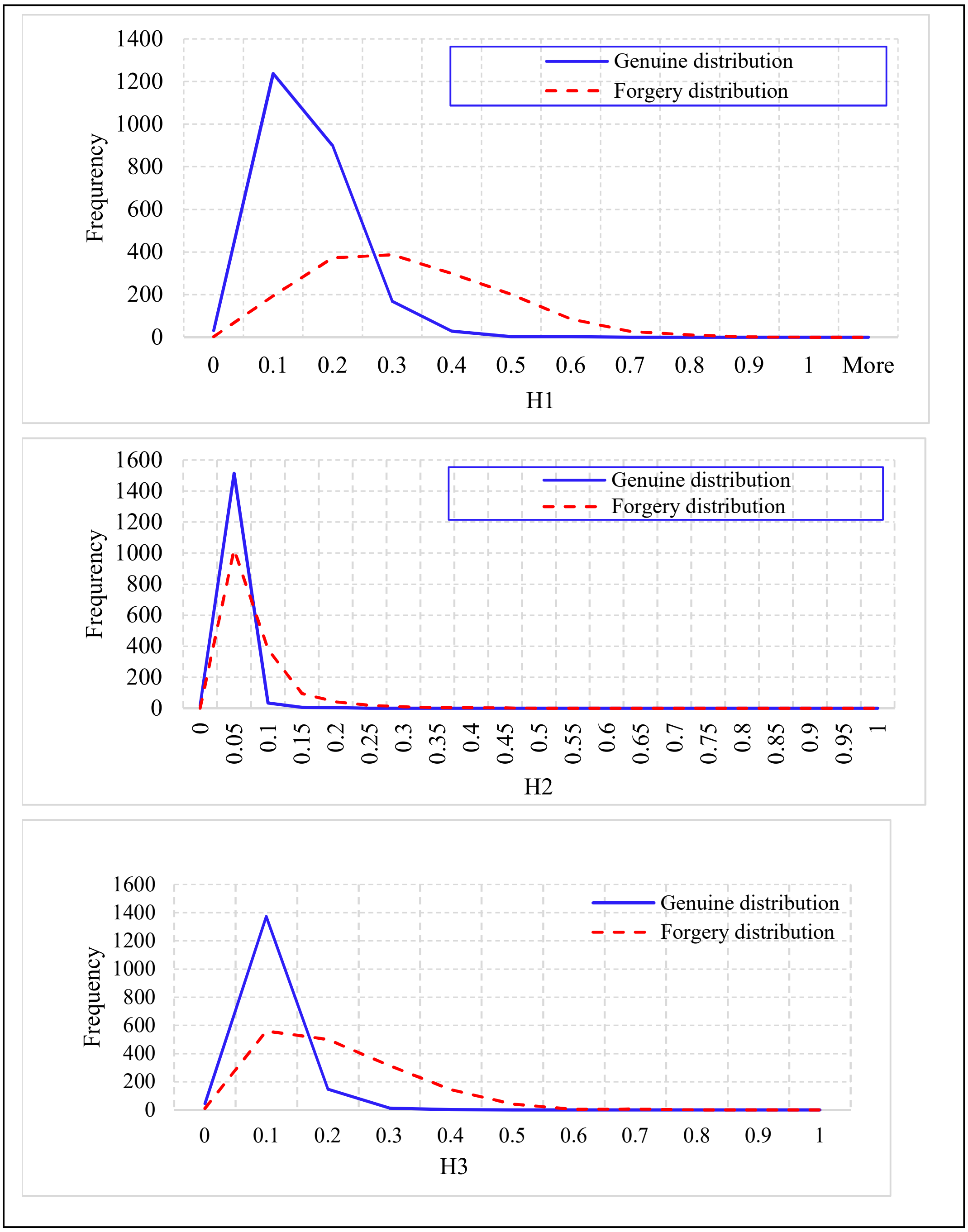


Figure 6.6. Frequency histograms of three objectively defined hesitation features (H1, H2, H3)

To assess whether these hesitation features are reliable as discriminators in signature analysis, consistency factors for each of the features have been analysed, as reported in [278]. Features associated with genuine signatures should be close to each other in value while distances between features associated with genuine and forged signatures should be relatively large. This discriminative capability of a feature is usually called the *consistency* of the feature and it is computed based on the statistics of the intra class (for the genuine signature class) and interclass (between the genuine and forged signature classes) distances. The consistency feature CF of a feature (adapted from [278] is calculated as defined in (6.4).

$$\mathrm{CF} = \frac{\mu_D(C_g\ ,C_g\ )-\mu_D(\ C_g\ ,C_f\ )}{\sqrt{\sigma_D^2(\ C_g\ ,C_g)}\ +\sqrt{\sigma_D^2(C_g\ ,C_f\ )}} \qquad \mathbf{(6.4)}$$

where Cg and Cf stand for the genuine and the forged classes, respectively, and where $\mu_D(C_g\ ,\ C_g\ )$ and $\sigma_D^2(\ C_g\ ,C_g)$; and $\mu_D(\ C_g\ ,\ C_f\ )$, and $\sigma_D^2(C_g\ ,C_f\ )$ are the sample means and sample variances of the genuine intra-class distances and the genuine-forged interclass distances, respectively.

To compute the consistency factor (CF) recalling the processing chain illustrated in Figure 2.1 (in Chapter 2), in the *feature extraction* step, a total of 60 features , which the literature shows are those most commonly used in signature processing [153], [209], [210], [275], [315], as defined in Table 2.1, 2.2 and 2.3 (in Chapter 2), are extracted from all samples of BioKent database described previously.

Subsequently, the *feature normalisation* step is carried out and the extracted features are normalised by using the MVN (Mean and Variance Normalisation) technique as defined in Equation 2.1 (in Chapter 2). Then the Spearman's rank correlation is evaluated between all the extracted and normalised features as explained in Section 2.1.2 (in Chapter 2) in the *feature correlation* step. As a result of this correlation test, 60*60 (3600) correlation values (ρ) are obtained.

Table 6.1. Uncorrelated features and their correlation values in the BioKent database

		Feature correlation values																							
		1	**2**	**3**	**4**	**5**	**8**	**10**	**13**	**14**	**17**	**19**	**22**	**23**	**24**	**26**	**27**	**28**	**29**	**34**	**37**	**46**	**50**	**59**	**60**
Feature correlation values	**1**	1.0	0.6	0.1	-0.3	0.1	0.6	0.2	-0.1	0.1	0.4	0.3	0.2	-0.1	-0.2	0.4	0.4	0.3	0.2	0.2	0.2	-0.1	0.4	-0.1	0.0
	2		1.0	0.5	-0.5	0.1	0.0	-0.3	-0.1	0.1	0.1	0.0	0.2	-0.2	-0.1	0.6	0.4	0.2	-0.1	0.5	-0.1	0.3	0.7	-0.1	-0.1
	3			1.0	-0.3	-0.1	-0.1	-0.2	-0.2	-0.1	0.2	0.2	0.2	-0.1	-0.1	0.3	0.2	0.2	-0.2	0.2	0.0	0.3	0.3	0.1	0.1
	4				1.0	-0.1	0.0	0.2	0.1	-0.1	-0.1	0.1	-0.1	0.2	0.3	-0.3	0.2	-0.6	-0.2	-0.3	0.0	0.1	-0.3	0.1	0.1
	5					1.0	-0.1	-0.1	0.0	0.1	-0.1	-0.1	0.0	0.0	-0.1	0.1	0.1	0.3	-0.2	0.1	-0.5	0.1	0.1	0.0	0.0
	8						1.0	0.3	0.1	0.1	0.6	0.3	0.1	-0.1	-0.1	0.0	0.1	0.2	0.3	-0.1	0.2	-0.3	-0.1	0.0	0.0
	10							1.0	0.0	0.0	0.1	0.5	-0.2	0.1	0.1	-0.2	-0.2	-0.1	0.2	-0.2	0.3	-0.6	-0.3	0.0	0.2
	13								1.0	0.3	0.3	0.0	-0.1	0.0	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	-0.2	-0.1
	14									1.0	0.0	0.1	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	0.1	-0.1	-0.1	0.0	-0.1	-0.3
	17										1.0	0.4	0.2	-0.2	-0.1	0.1	0.1	0.2	0.1	0.0	0.1	-0.1	0.1	0.0	0.0
	19											1.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.2	-0.3	0.0	0.0	0.1
	22												1.0	-0.4	0.0	0.2	0.3	0.1	0.1	0.0	0.0	0.3	0.1	0.0	0.0
	23													1.0	0.3	-0.1	-0.1	-0.1	0.0	-0.2	0.1	-0.1	-0.1	0.0	0.0
	24														1.0	0.0	0.2	-0.2	0.0	-0.1	0.0	0.1	-0.1	0.0	0.0
	26															1.0	0.3	0.1	-0.1	0.3	-0.1	0.2	0.4	0.0	-0.1
	27																1.0	0.0	-0.1	0.2	-0.1	0.5	0.5	0.0	0.0
	28																	1.0	0.1	0.1	-0.1	-0.1	0.1	0.0	0.0
	29																		1.0	-0.1	0.4	-0.2	0.0	0.1	0.0
	34																			1.0	-0.1	0.2	0.3	0.0	-0.1
	37																				1.0	-0.4	0.0	0.0	0.1
	46																					1.0	0.5	0.0	0.0
	50																						1.0	0.0	0.0
	59																							1.0	0.3
	60																								1.0

The correlation results obtained in the experiment are illustrated in Table 6.1 and the uncorrelated features are listed in Table 6.2.

Table 6.2. Uncorrelated features and three hesitation features

Feature Number	Feature Names
1	Total distance of pen travelled
2	Total signature execution time
3	Pen lift:(Number of pen ups=> button 1 to 0)
4	Average velocity in X direction
5	Average velocity in Y direction
8	Maximum pen velocity in x - Average pen velocity in x
10	Maximum pen velocity in y - Average pen velocity in y
13	Average pen acceleration in x
14	Average pen acceleration in y
17	Maximum pen acceleration in x - Average pen acceleration in x
19	Maximum pen acceleration in y - Average pen acceleration in y
22	Azimuth
23	Altitude
24	Pressure
26	Sum of x coordinate values
27	Standard deviation x coordinate values
28	Maximum x coordinate value - Last x coordinate value
29	First x coordinate value - minimum x coordinate value
34	Sum of y coordinate values
37	First y coordinate value - minimum y coordinate value
46	width/height ratio
50	Number of vertical midpoint crossing the signature
59	Average pen jerk in x
60	Average pen jerk in y
61	H1
62	H2
63	H3

Following the determination of the uncorrelated features, the consistency factor is calculated as defined in (6.4), including the three hesitation features with the 24 uncorrelated features (Table 6.2). A comparison of the consistency factors of different features is shown in figure 6.7 using boxplots for 79 users in a non-increasing manner showing three hesitation features (61, 62 and 63) among the best four features on the left. This high consistency of the objectively defined hesitation features indicates that these could be useful in automatic handwritten signature verification.

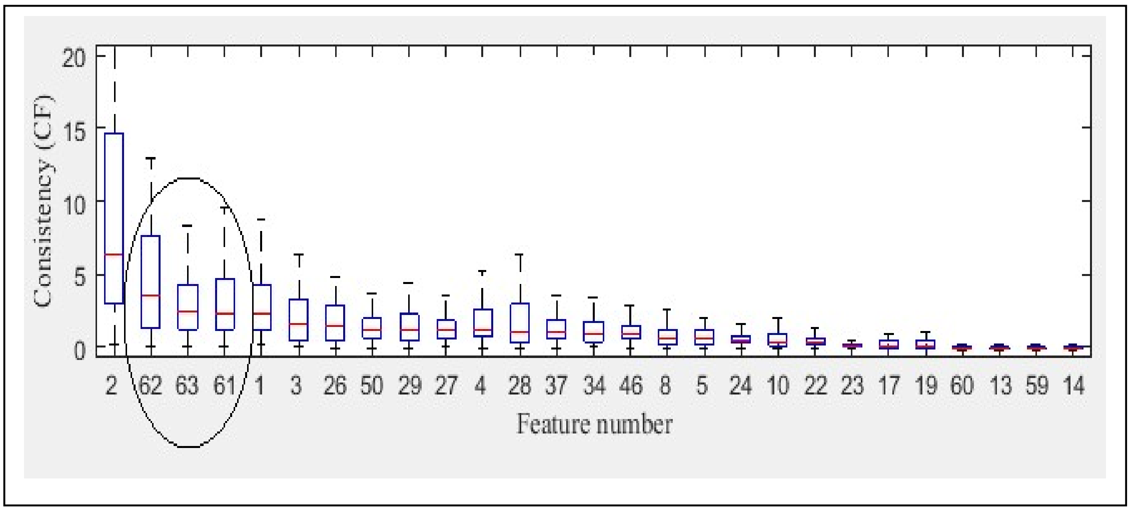


Figure 6.7. Consistency factor of all the features

Also, to learn more about the importance of these features, a Gain Ratio Attribute Evaluation (which evaluates the importance of an attribute by measuring the gain ratio with respect to the class) and a Correlation Attribute Evaluation (which evaluates the importance of an attribute by measuring the Pearson's correlation between it and the class) feature selection method have been implemented using the WEKA package [220], for which a higher gain ratio (in Gain Ratio Attribute Evaluation) and a higher correlation coefficient (in Correlation Attribute Evaluation) for a feature indicates better discriminative power for classification

[316]. H1, H2 and H3 are added to the feature list one at a time (i.e. ranking is performed for the 24 uncorrelated features + either H1 or H2 or H3). For the sake of notation, the different combinations (including these three combinations and some others which are utilised later in this section) are denoted as shown in Table 6.3. The resulting ranking in Table 6.4 (for SetH1, SetH2 and SetH3) shows that H1 (feature 61) and H3 (feature 63) are at the top of the ranking in SetH1 and SetH3 respectively in both feature selection methods, and H2 (feature 62) is at rank two and rank three in SetH2 in Gain Ration and Correlation Attribute Evaluation methods respectively. It is very clear from these ranking experiments that all these three objectively defined hesitation features are important features, showing a high discriminative power for classification.

Table 6.3. Feature combination sets

Feature sets	Feature combinations
SetH0	Only 24 uncorrelated features
SetH1	All 24 uncorrelated features & H1
SetH2	All 24 uncorrelated features & H2
SetH3	All 24 uncorrelated features & H3
SetH1+H2+H3	All 24 uncorrelated features & H1 , H2 , H3
SetH1+H2	All 24 uncorrelated features & H1& H2
SetH2+H3	All 24 uncorrelated features & H2 & H3
Set H1+H3	All 24 uncorrelated features & H1 & H3

Table 6.4. Evaluation of features using Gain ratio and Correlation based feature selection algorithm

Feature Ranking	Gain ratio				Correlation based		
	SetH1	SetH2	SetH3		SetH1	SetH2	SetH3
1	**61**	2	**63**		**61**	2	**63**
2	2	**62**	2		2	26	2
3	26	26	26		26	**62**	26
4	50	50	50		34	34	34
5	34	34	34		4	4	4
6	4	4	4		1	1	1
7	10	10	10		27	27	27
8	17	17	17		10	10	10
9	1	1	1		8	8	8
10	19	19	19		37	37	37
11	27	27	27		46	46	46
12	8	8	8		24	24	24
13	5	5	5		5	5	5
14	13	13	13		19	19	19
15	37	37	37		17	17	17
16	46	46	46		29	29	29
17	14	14	14		22	22	22
18	59	59	59		13	13	13
19	60	60	60		50	50	50
20	22	22	22		28	28	28
21	29	29	29		60	60	60
22	28	28	28		14	14	14
23	24	24	24		3	3	3
24	3	3	3		23	23	23
25	23	23	23		59	59	59

To assess the viability of these features in identifying genuine and forged signature samples, classification is performed by using a simple KNN classifier (K=1) using a 10-fold cross validation methodology. The Weka [220] software is used with default settings in order to do this. Eight different combinations of feature sets, as shown in Table 6.4, are used to compare the performance of classification in different sets. Classification accuracy and error rates are evaluated for each set and the results are presented in Figure 6.8. The best classification performance is achieved (the lowest error rate) using feature combination SetH1 and the highest error rate using SetH0. The difference in error rates is not large here, but it is evident that whatever measure of hesitation we take (H1, H2 or H3), using feature sets which include one or two or all of these hesitation features increases the accuracy of classification and reduces the error rates. Also, it is noticeable that the two measures of hesitation adopting a time-based representation show comparatively better results than the distance based representation, and this is also supported by the consistency factors and feature ranking (using Gain Ratio and Correlation) outcome.

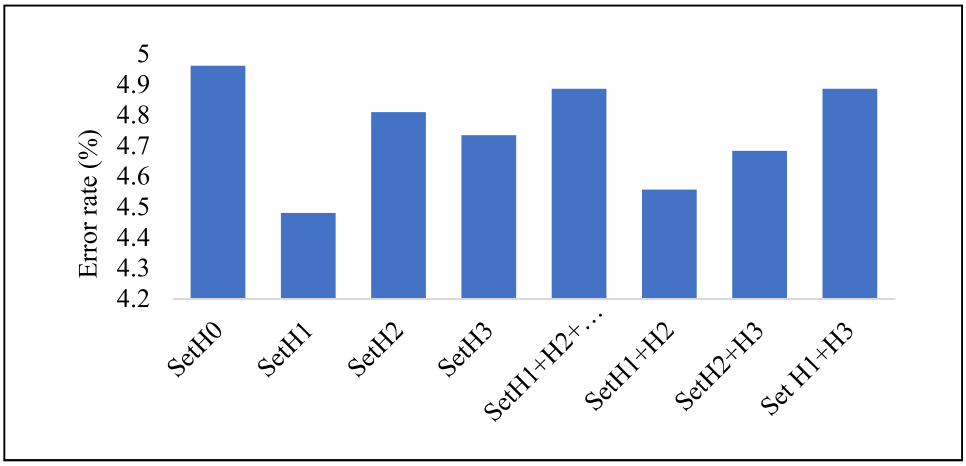


Figure 6.8. Classification results with different feature sets

6.4 Conclusions

In this chapter, initially a brief review of the notions of fluency and hesitation in handwriting/signature production has been provided, where it has been shown that the hesitation phenomena (as a lack of fluency) has been used in different research areas (including children's handwriting development, psychology, speech analysis, neuroscience, forensic document examination, and biometrics) but generally more qualitatively than quantitatively. So subsequently, three quantitative approaches have been defined and described. All the definitions are simple and intuitive, and all are easily measurable objectively and algorithmically. Thus, the feature extraction process is rapid, reliable and robust. Using these definitions features have been extracted from the signature samples in the RevKent and BioKent databases. Influences of these extracted hesitation features on creating new signature (naturally revoked) and recognising genuine and forged signatures have been explored revealing that signers are more hesitant initially in signing the new signatures than the original, as might be expected, and gradually the hesitancies reduce with time, showing signers' confidence in signing the new signature as time progress. In a genuine and forgery signature analysis scenario, the consistency factors have been evaluated for each feature, revealing high values for these hesitation features. The outcomes of two feature ranking analyses have also shown that these features have good discriminative power, and results from a classification process show an increase in accuracy in discriminating between genuine and imitated signatures when a hesitation feature is included in the feature set.

Of course, there are many further issues which need to be investigated, such as correlations between the hesitation feature and type of signature, its complexity and so on, but this initial study has established some basic indicators of performance and the potential for further development of such features in automatic signature verification. The study has also provided further insights into the adoption of the idea of natural revocability in the context of handwritten signatures. The experimental analysis presented in this chapter shows that even very simple and intuitive measures of hesitation offer an encouraging and effective performance improvement, which suggests that developing more refined and more powerful measures may be worthwhile. From the point of view of the forensic analysis of signatures, this work suggests that developing better and more objective ways of assessing hesitation in the signing process may have an increasing role to play in forensic investigations, especially as an increasing trend towards online capture of signature data continues.

The next chapter will present a summary of all the studies performed and the contributions made by the work reported in this thesis, and also will identify further questions and further research ideas which should be seen as priorities for future investigation.

Chapter 7:

Some ancillary issues and final remarks

This chapter will present a number of additional pieces of experimental and analytical work carried out to help to complete a comprehensive picture of various issues relevant to the study reported in the main body of the thesis, as well as a final discussion of all the contributions made., This will link naturally to a brief discussion of possible future work required to develop improved strategies for handwritten signature biometric systems in the light of the findings emerging from the overall study.

The chapter will start in Section 7.1 with a discussion on some initial additional work that can be seen as a bridge between the work reported so far and some areas on which future work might most usefully be focused. Section 7.2 will summarise the work reported in this thesis, while pointing to some priority areas for future work in Section 7.3. Finally, Section 7.4 will conclude the chapter.

7.1 Some ancillary issues

In addition to the main study presented in detail in the previous chapters, this section reports - briefly rather than in the detail provided on the principal experiments described earlier - on some additional work which has been carried out, but not yet fully developed. This will enhance the main body of the work described earlier, increasing the value of the study overall and creating a platform for possible future work. Some of these additional initial works are presented as follows:

7.1.1 Human performance in the context of natural revocability

To investigate human perceptual capabilities with respect to differences between a longstanding original signature and a newly created signature, thereby shedding further light on the natural revocation process, the following experiments have been carried out.

- **Forgery detection**

Initially, five signatures have been chosen from the original signatures of the Rev-Kent database (described in Chapter 3) which are seen to represent a variety of styles, designated as 'O-target' signatures, and five new signatures of these same signers have also been chosen, designated as 'N-target' signatures, and then these signatures were presented to a group of five different subjects (designated as 'forgers'). These five subjects were mainly university students. For each target signature (both O-target and N-target samples) the forgers were required to submit five forgery samples for each target signature. The subjects (forgers) were allowed to practise each target signature on a piece of paper before imitating on the graphics tablet, and to keep the respective target signature in view during the imitation process. They could delete and resubmit samples when they felt they had made a mistake, to ensure that mistakes made due to lack of familiarisation with the digitising tablet is narrowed down to a minimum.

Another twenty-five participants, mainly university staff and students together with members of the general public of different professions (e.g. banker, web developer, businessman, detective etc.) took part in this experiment. A range of both genuine and forgery signature samples (designated as 'test signatures') for the original and the new signatures were presented to the participants and they were asked to classify each as either genuine or forgery, and they were also asked to rate each on a scale of 1 to 10, reflecting how likely it was considered that the test signature sample was a genuine sample (where 1 corresponds to definitely genuine and 10 corresponds to definitely forgery), with respect to five genuine samples (designated 'reference samples') for each target signature, which was simultaneously in view. For each target signature, five genuine test samples and five forgery test samples of the target signature were displayed in a one-by-one fashion, while the five genuine reference samples of the target were constantly in view. A total of one hundred test signature samples, where fifty of them were original and the other fifty were new signatures, were presented to the participants.

In addition to classifying the test samples as described above, the participants were asked to comment on the factors or signature characteristics that they considered important when rating the test sample as genuine or forged, and this is discussed further later in this section.

The performance achieved from this initial experiment in human visual inspection of the genuine and imitated/forged samples of handwritten signatures is presented in Table 7.1. These results show that an average of 83.60% of the original test signatures and 86.96% of the new test signatures have been correctly classified as being either genuine or forged; and 16.40% and 13.04% of the original and new signatures (respectively) have been erroneously classified. 20.64% of the forgery test samples have been falsely accepted as genuine (FAR) and 12.16% of the genuine signature samples have been falsely rejected as forgeries (FRR) in the original signature; and in new signature false acceptance (FAR) and false rejection (FRR) rates are 13.12% of 13.60% respectively. The FAR is calculated as the percentage of forgeries falsely being accepted out of the total number of forgeries shown, and similarly the FRR is calculated as the percentage of falsely rejected genuine signatures out of the total number of genuine signatures shown.

Table 7.1. Average human performance

Average Human Performance		
	Original	**New**
Total correct classification %	83.60	86.96
Total error %	16.40	13.04
FAR %	20.64	13.12
FRR %	12.16	13.60

Overall, the performance of human verification of the handwritten signatures is promising in both the original and newly generated signatures, with marginally better performance achieved for the new signatures. Although, as noted earlier, this experiment should be regarded as very much at a very preliminary stage of potential a much broader and more detailed investigation, it suggests that the possibility of using a new/replacement signature being in signature verification in circumstances where the longstanding original signature is compromised, is a viable option, and supports optimism about the practical potential of our proposed concept of the value of natural revocability. This also provides a platform for future work in this area.

- **Similarity analysis**

A similar experiment is also conducted to analyse the similarity scores for the five target signatures in original and new signatures described in the above experiment. In this experiment, twenty-five participants were presented with five genuine samples of each target signature (both original and new) and asked to rank each test signature sample according to its visual similarity level (on a scale of 1 to 10, where 1 corresponds to the most similar and 10 corresponds to the least similar) with respect to the reference signature samples for each target signature. The frequency histograms of the responses are shown in Figure 7.1 and Figure 7.2.

It can be seen from Figure 7.1 and Figure 7.2 that the similarity score histograms show a very similar pattern for the target signature in the original and new signatures, indicating that the observers perceived the similarity of the signature samples with the reference samples provided in a similar manner; in other words, the subjects could see the similarity in the new signatures (which have not been subject to the same degree of long-standing practice and repetition) as they could observe in the much more established original signatures (i.e. visually the difference between samples of a long-time practised signature and a rather short-time practised signature is not noticeable) .

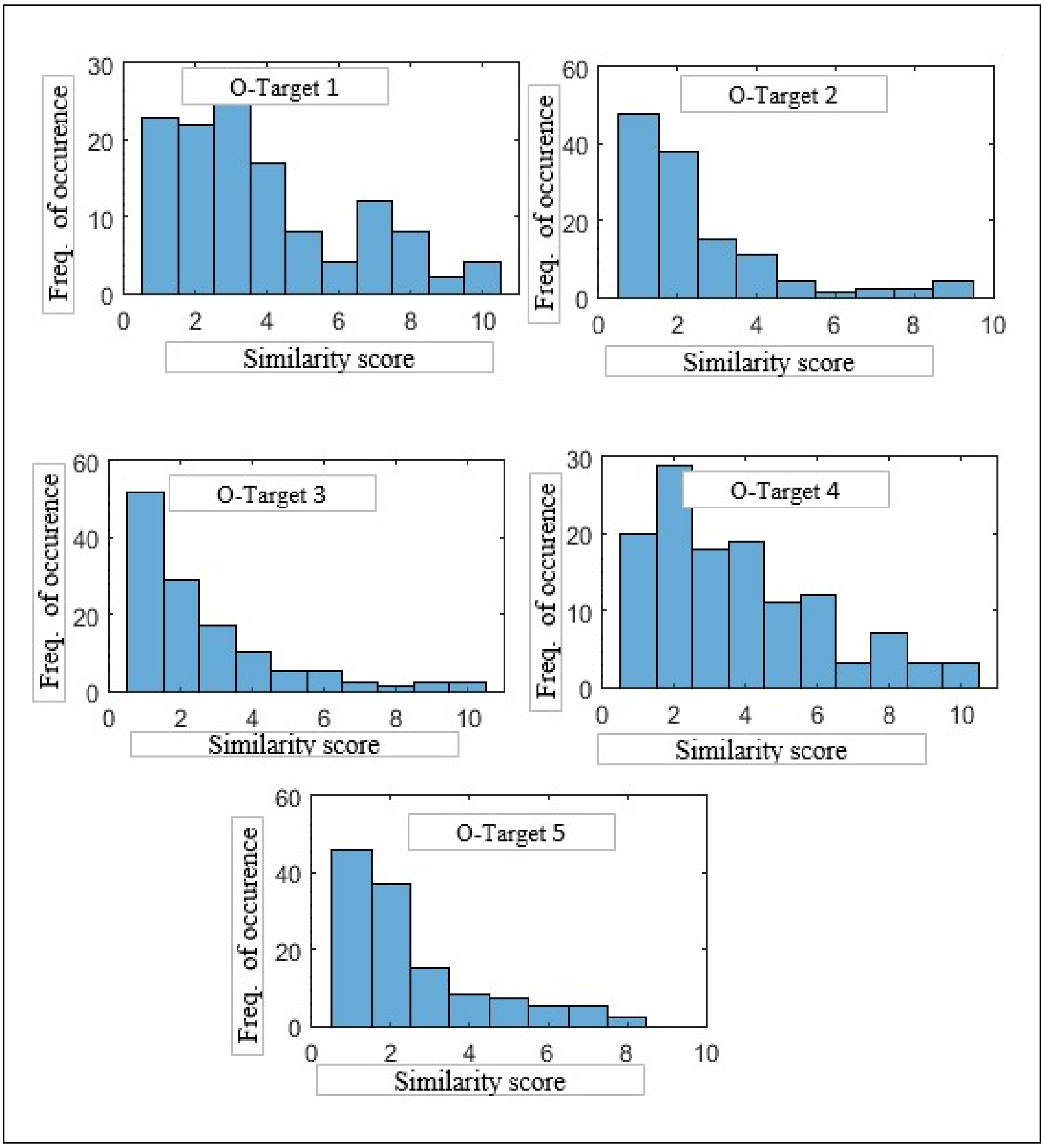


Figure 7.1. Frequency histogram of similarity scores in new signatures

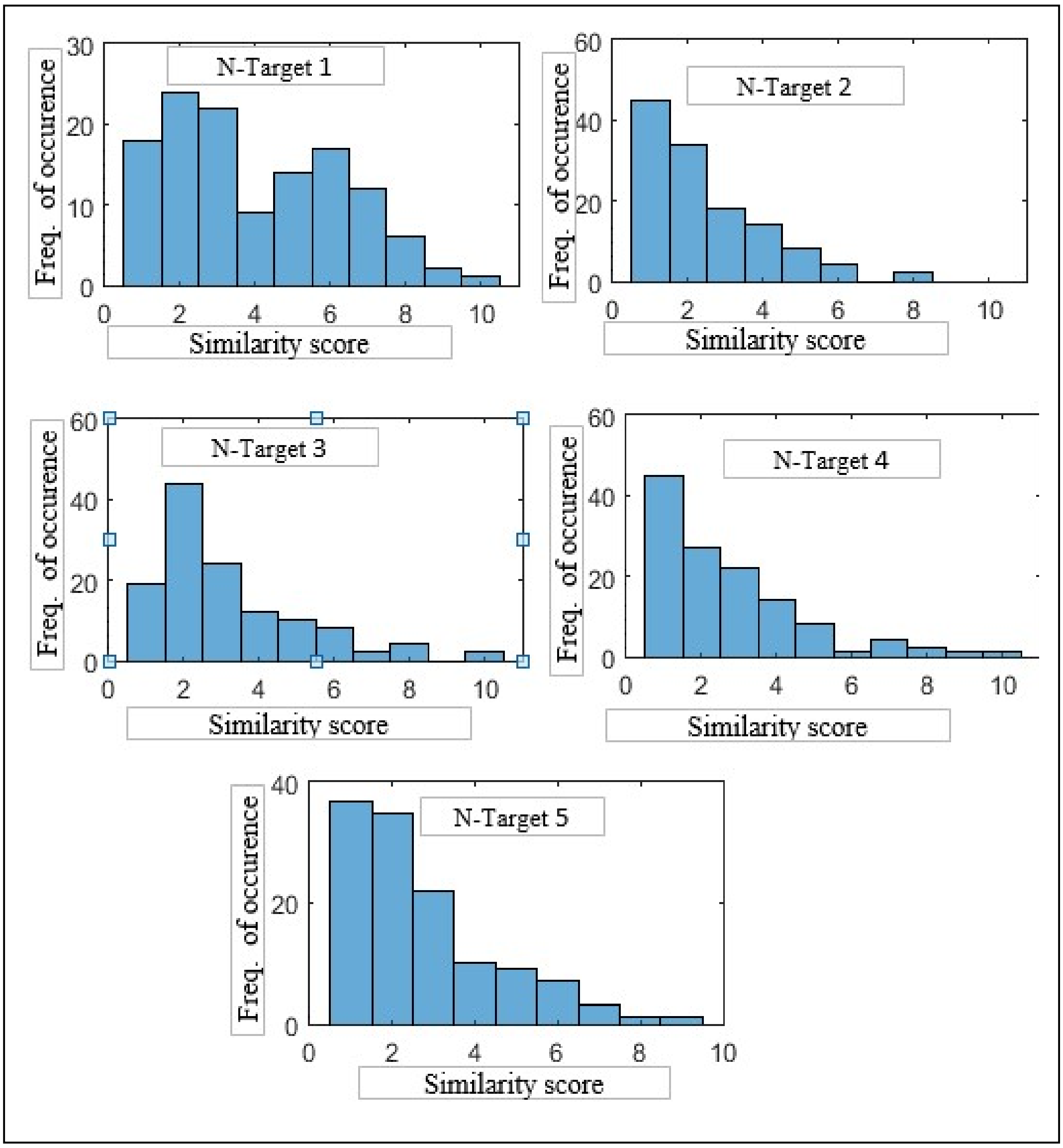


Figure 7.2. Frequency histogram of similarity scores in new signatures.

- **Analysis of participant comments**

Analysing the comments of the participants (received during the experiment carried out for forgery detection, described earlier in this section), it is found that participants particularly took note of overall signature style, start and/or end of the signature, size of letters in relation to other letters, fluency, smoothness of the line etc. to decide whether the test sample is a genuine or forged one. Among these perceptual characteristics, fluency or disfluency in the line was mentioned by a majority of the participants to have played the most important role in their decision; which is also one of the primary features important to the forensic experts in distinguishing genuine and forgery signatures (described in Chapter 6). Most of the participants mentioned that a lack of fluency (or hesitation) or lack of smoothness in the line of the signature suggested the test sample being forged, as quoted from one of the participants' comments, as follows: "…with some there appeared to be hesitation in the line, making it less fluid and bumpy suggesting a copy was being made". Signature style, shape and size were also mentioned by some of the participants as influential criteria, where they looked at the shape, size and structure of individual characters in the signature, beginning and ending parts of the signature, slant of letters, spacing between characters etc. This preliminary analysis of the comments shows that fluency or hesitation, which is an important key feature in signature authentication by forensic experts, also plays an important role in distinguishing between genuine and forged samples by non-experts as the participants in this experiment. This vindicates the work on developing objective algorithmic representations of hesitation described in the main body of the thesis, and also suggests that a further investigation, developing more refined and more powerful measures of the hesitation feature (as described in Chapter 6) may be worthwhile not only in forensic investigations but also in general automatic signature verification.

7.2 Summary of the work carried out

The study reported in this thesis represents a comprehensive analysis of the particular characteristics of a behavioural modality. Specifically, the handwritten signature has been investigated as the target biometric modality of interest by means of an experimental study using a collection of signature samples, and wide-ranging subsequent analysis.

An overview of the general biometric system, its security concerns and protection mechanisms currently relevant to the study reported in this thesis is presented initially in Chapter 1. Several protection mechanisms (in particular, raising the notion of revocable biometrics) for different biometric modalities are discussed, and it was noted from a study of the literature that the concept of natural revocability, (the fact that most behavioural biometrics, being under the direct control of the "user", can be created at will in multiple forms) has not been studied in relation to biometrics and data revocability until the study reported in this thesis in Chapter 4 (also published in [152]).

Following this, the advantages of the handwritten signature over other biometrics for use in applications such as a point-of-sale environment has been discussed (Section 1.4). The complex process of signing and signature development is further analysed. Several factors affecting the inherent variability of the signature (e.g. physical and mental condition of the writer; the country, culture or the school where the writer has been taught etc.) have been described. A review of signature-based authentication systems reported in the literature has been presented with their application in different disciplines as well as biometrics, identifying some important issues in relation to objective and reliable methods for handwritten signature analysis.

The literature review discussed in Chapter 1 provided a platform for the development of a set of principal objectives to be reported as the core of this thesis.

Thus, in this study the focus has been to analyse and explore two of the very closely related areas referred to in the initial review in Chapter 1, which are: first, the possibilities of adopting a natural revocability strategy in relation to security and reliability in handwritten signature analysis; and second, the development of features which may be particularly effective in the area of handwriting analysis.

Before any analysis, an overview of the basic experimental infrastructure and the important practical details used in all the experiments reported in this thesis has been provided in Chapter 2. This is to make the subsequent chapters cohesive and easier to follow. Thus, the details of data acquisition, feature extraction and some feature processing (respectively), and classification techniques implemented in the experiments and analysis for the study reported in this thesis have been presented. A review of some of the principal online signature databases available to researchers has been presented which provided a useful and critical analysis of the potential additional benefits and enhanced characteristics of the proposed new data acquisition exercise (described in Chapter 3). The deficiencies in the publicly available databases in relation to the requirements of the proposed study made the compilation of a new and "bespoke" database a fundamental part of the work in the study of interest here. The two databases utilised for the experiments carried out in the subsequent chapters have also been identified in this Chapter. This data collection exercise was extremely time-consuming (and not always easy to accomplish), the benefits of doing this are self-evident.

The data collection protocol is discussed in Chapter 3 to define the rigorous methods used to develop the strategy for the uniform collection of the signature samples. A review of the ethical procedures is presented detailing the criteria to be adhered to for participation in the data collection. An overview of the acquisition system, both hardware and software, together with an illustration of the collected data and subject information has been provided. The challenges during data acquisition procedure and the importance of this entirely new database collected as part of the study have been discussed. After the overall consideration of the basic tools and experimental infrastructure in Chapter 2 and the new database collected for this study described in Chapter 3, the subsequent chapters have reported and analysed a range of specific experimental studies to address the main aims of the overall study.

A further review of the revocability strategies in different biometric modalities is provided and discussed in Chapter 4, after the initial review provided in Chapter 1, together with a detailed explanation of the idea of natural revocability in behavioural biometrics, specifically in handwritten signature biometrics. Subsequently, the suitability and effectiveness of natural revocability in handwritten signature biometric as a practical option in signature recognition, have been investigated by observing how "stability" of the form of the signature changes over a period of time, as the stability in signing is a key factor in determining the suitability of the signature for biometric identification. Initially the stability has been analysed for the signature samples collected in four successive capture sessions and later for ten successive captures sessions, and the analysis suggested that if a sufficient time period is allowed then there is a high likelihood of convergence in stability between a highly practised and long-standing signature and an alternative new representation.

A further investigation of the characteristics of potential revocability in the signature modality has been carried out by analysing performance invoking the "biometric menagerie" notation for individual behavior, which was first introduced by Doddington et al. in the context of speaker recognition [254]. Results from this analysis provide initial data to show how stability patterns between original and new signatures are likely to change across individuals and suggests that it is possible nevertheless to achieve stability when changing their signature styles. Recognition performance has also been evaluated as a more practically-oriented test of the viability of the natural revocability concept and has shown that the new signatures can be reliably recognised, and even that the performance achieved can be better than that found for the original signatures.

A feature-based analysis of the natural revocability concept has been presented in Chapter 5 by investigating some features commonly used in signature processing in both original and new signatures of a group of writers, and exploring the relationship between features, signature styles and their effects in relation to original signatures and new signatures. The results from this analysis have shown that some features such as pen inclination (azimuth, altitude etc.), pen pressure, velocity, acceleration related features remain almost the same for most of the signers, when they change their signatures to create a new one, suggesting that the same underlying constructional mechanism is evident in the new signatures as was the case with the original signatures for these features. Some other features have been found to be different between the original and new signatures, such as number of times pen is lifted, total time and distance travelled to execute the signatures, and so on. Intuitively, the difference in size, shape or the style between the original and new signatures possibly contributed to the difference and this has been investigated by an experimental analysis. Feature variability in original and new signatures has been analysed and it has been found that the ranking of the features based on the variability is almost identical in original and new signatures (though the actual variability is not the same but very similar), indicating the intrinsic property of the handwritten signature (i.e. intra-person variability) remains consistent when developing new signatures through the natural revocation process. A further analysis has been carried out to investigate the consistency of the signature style and effect of style and other factors (e.g. age, gender, and handedness) on the features in the original and new signatures.

To provide further insights into the adoption of the idea of natural revocability in the context of handwritten signatures, development of a type of feature related to the concept of hesitancy (or its converse, fluency) is explored in Chapter 6. A brief review of the notions of fluency and hesitation in handwriting/signature production provided initially in this chapter has shown that hesitation has been important in different research areas (including children's handwriting development, psychology, speech analysis, neuroscience, forensic document examination, and biometrics) but generally more qualitatively than quantitatively. So subsequently, three simple and intuitive quantitative approaches have been defined and described. All of these definitions are easily implementable objectively and algorithmically, making the feature extraction process rapid, reliable and robust. Exploring the influences of these objectively defined hesitation features in creating the new signature has revealed that signers are more hesitant initially in signing the new signatures than the original - as might be expected - and gradually the hesitancies reduce with time, showing signers' increasing confidence in signing new signatures as time progresses; and investigating recognising genuine and forged signatures showed that the hesitation is higher in forgery signatures than in genuine signatures, agreeing with the qualitative definition applied in a typical forensic scenario. Also, evaluation of consistency factors for each feature, including these hesitation features, has revealed high consistency values for these features in genuine and forgery signature analysis scenarios. The consistency factors have been evaluated for each feature, revealing high values for these hesitation features and outcomes of two feature ranking analyses have also shown that these features have good discriminative power. Results from a classification process show an increase in accuracy in discriminating between genuine and imitated signatures when a hesitation feature is included in the feature set. The study reported in Chapter 6 has shown that even very simple and intuitive measures of hesitation offer an encouraging and effective performance improvement, suggesting that developing more refined and more powerful measures may be worthwhile in the future.

A preliminary investigation is conducted in this final chapter (Chapter 7) to analyse the concept of natural revocability in a signature verification experiment based on human visual inspection. Although this investigation has not been carried out in as detailed or rigorous a way as the main core of the study reported in this thesis, it has given some indication that the natural revocability concept could be suitable to be exploited in signature verification and more generally signature authentication (along with the study reported in Chapter 4). An initial similarity analysis by human visual inspection of the new and original signatures also suggests that the new signature can be similarly detected if the original signature is compromised. This preliminary study has provided some further insight into the development of the hesitation features (along with the study reported in Chapter 6).

The study reported in this thesis has afforded us the opportunity to investigate issues related to security and reliability in biometrics, especially using the handwritten signature as a wider biometric modality by exploring two closely related areas in handwritten signature. It has also enabled us to develop significant new work and contribute to this general area, with the aim of providing some new insights on relevant issues of importance.

7.3 Future work

Some possible new research ideas emerged from the study reported in this thesis are as follows:

- Although the data collection process described in Chapter 3 of this thesis has been very challenging and time consuming, a larger dataset with participants from a wider range of ages, occupations etc. would greatly benefit further investigation into natural revocability analysis in the future. For instance, investigating whether people from a particular age group or occupation (e.g. people who have to sign on a daily basis in their occupation or who do not sign very frequently) attain stability quicker/slower than others or not.

- Considering the preliminary work reported in Section 7.1 in this chapter, a larger number of participants to include in the human visual inspection may be suggested for further detailed investigation.

- As the original signature, the new signature also can be forged, and thus an investigation of the resistance to forgery of the new signatures compared to the original signature may be suggested. This can be performed as an extension of the forgery data collection described in Section 7.1 in this chapter, introducing a ranking based on the ‘difficulty to forge’ for a larger number of signatures to be forged by more participants (forgers).

- An investigation of the correlation between each of the hesitation features as described in Chapter 6 and the type of signature, its complexity, and so on, may be useful in deciding on which hesitation feature to use for what type of signature.

7.4 Conclusion

This chapter has presented a summary of the research studies performed and significant contributions made to the field of handwritten signature biometrics by the work reported in this thesis, together with some initial findings leading to some possible future work priorities for further investigation.

The work reported in this thesis is hoped to have a positive influence on issues related to security and reliability in handwritten signature biometrics.

Bibliography

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

[2] A. K. Jain, K. Nandakumar, and A. Ross, *50 years of biometric research: Accomplishments, challenges, and opportunities*, vol. 79. 2016, pp. 80–105.

[3] J. Ortega-Garcia *et al.*, 'The multiscenario multienvironment BioSecure Multimodal Database (BMDB).', *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 6, pp. 1097–111, Jun. 2010.

[4] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, *Guide to biometrics*. Springer, 2004.

[5] R. D. Labati, A. Genovese, E. Muñoz, V. Piuri, F. Scotti, and G. Sforza, 'Biometric Recognition in Automated Border Control', *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–39, Jun. 2016.

[6] A. K. Jain, A. Ross, and S. Prabhakar, 'An Introduction to Biometric Recognition', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[7] D. Meuwly and R. Veldhuis, 'Forensic biometrics: From two communities to one discipline', *of the International Conference of the*, 2012.

[8] A. K. Jain, A. Ross, and U. Uludag, 'Biometric Template Security : Challenges and Solutions', *Security and Watermarking of Multimedia*, vol. 4675, no. IV, pp. 629–640, 2002.

[9] A. K. Jain, K. Nandakumar, and A. Nagar, 'Biometric Template Security', *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, p. 579416, 2008.

[10] F. Hillerstrom, A. Kumar, and R. Veldhuis, 'Biometrics Special Interest Group (BIOSIG), 2014 International Conference of the', *Biometrics Special Interest Group (BIOSIG), 2014 International Conference of the*. pp. 1–9, 2014.

[11] D. Sadhya, S. K. Singh, and B. Chakraborty, 'Review of key-binding-based biometric data protection schemes', *IET Biometrics*, vol. 5, no. 4, pp. 263–275, 2016.

[12] J. Hermans, R. Peeters, and B. Mennink, 'Biometrics Special Interest Group (BIOSIG), 2014 International Conference of the', *Biometrics Special Interest Group (BIOSIG), 2014 International Conference of the*. pp. 1–6, 2014.

[13] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle, 'Generating cancelable fingerprint templates', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 561–572, Apr. 2007.

[14] A. M. P. Canuto, F. Pintro, A. F. Neto, and M. C. Fairhurst, 'Enhancing Performance of Cancellable Fingerprint Biometrics Using Classifier Ensembles', in *2010 Eleventh Brazilian Symposium on Neural Networks*, 2010, pp. 55–60.

[15] V. M. Patel, N. K. Ratha, and R. Chellappa, 'Cancelable Biometrics: A review', *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 54–65, Sep. 2015.

[16] Y.-L. Lai *et al.*, 'Cancellable iris template generation based on Indexing-First-One hashing', *Pattern Recognition*, vol. 64, pp. 105–117, 2017.

[17] R. Plamondon and N. Srihari, 'On-Line and Off-Line Handwriting Recognition : A Comprehensive Survey', vol. 22, no. 1, 2000.

[18] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, ‘Learning features for offline handwritten signature verification using deep convolutional neural networks’, *Pattern Recognition*, vol. 70, pp. 163–176, 2017.
[19] A. K. Jain, A. Ross, and S. Pankanti, ‘Biometrics: A Tool for Information Security’, *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, Jun. 2006.
[20] K. J. Hanna and H. T. Hoyos, ‘Methods for performing biometric recognition of a human eye and corroboration of same’, US Patent 9,613,281, 2017.
[21] A. K. Jain and A. Kumar, ‘Biometrics of Next Generation: An Overview’, 2010.
[22] A. Adler, ‘Biometric System Security’, *Handbook of Biometrics*, pp. 381–402, 2008.
[23] I. Chingovska, A. Anjos, and S. Marcel, ‘On the effectiveness of local binary patterns in face anti-spoofing’, *Biometrics Special Interest*, 2012.
[24] J. Galbally, J. Fierrez, J. Ortega-Garcia, and R. Cappelli, ‘Fingerprint Anti-spoofing in Biometric Systems’, 2014, pp. 35–64.
[25] S. Marcel, M. S. Nixon, and S. Z. Li, Eds., *Handbook of Biometric Anti-Spoofing*. London: Springer London, 2014.
[26] P. Reddy, A. Kumar, and S. Rahman, ‘A new antispoofing approach for biometric devices’, *IEEE Transactions on*, 2008.
[27] N. Ratha, J. Connell, R. M. Bolle, and S. Chikkerur, ‘Cancelable biometrics: A case study in fingerprints’, *Proceedings - International Conference on Pattern Recognition*, vol. 4, pp. 370–373, 2006.
[28] U. Uludag, A. Ross, and A. Jain, ‘Biometric template selection and update: a case study in fingerprints’, *Pattern Recognition*, vol. 37, no. 7, pp. 1533–1542, Jul. 2004.
[29] J. Feng and A. Jain, ‘FM model based fingerprint reconstruction from minutiae template’, *Advances in Biometrics*, 2009.
[30] A. Ross, J. Shah, and A. K. Jain, ‘From Template to Image: Reconstructing Fingerprints from Minutiae Points’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 544–560, Apr. 2007.
[31] E. Liu and K. Cao, ‘Minutiae Extraction From Level 1 Features of Fingerprint’, *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1893–1902, Sep. 2016.
[32] K. Cao and A. K. Jain, ‘Learning Fingerprint Reconstruction: From Minutiae to Image’, *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 104–117, Jan. 2015.
[33] A. Adler and S. A. C. Schuckers, ‘Biometric Vulnerabilities, Overview’, in *Encyclopedia of Biometrics*, Boston, MA: Springer US, 2015, pp. 271–279.
[34] C. Tilton and M. Young, ‘Study Report on Biometrics in E-Authentication (M1/07–0185). American National Standards Institute’, *International Committee for Information Technology*, 2007.
[35] F. Abdullayeva and Y. Imamverdiyev, ‘Analysis of security vulnerabilities in biometric systems’, *The second international conference: problems of cybernetics and informatics*, 2008.
[36] J. Cohn, ‘Keeping an eye on school security: The Iris Recognition Project in New Jersey schools’, *NIJ Journal*, 2006.
[37] U. B. W. Group, ‘Biometric security concerns’, 2003.
[38] S. Yoon, J. Feng, and A. Jain, ‘Altered fingerprints: Analysis and detection’, *IEEE transactions on pattern analysis*, 2012.

[39] A. Czajka, 'Pupil Dynamics for Iris Liveness Detection', *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 726–735, Apr. 2015.

[40] 'Magazine photos fool age-verification cameras ~ Pink Tentacle'. [Online]. Available: http://pinktentacle.com/2008/06/magazine-photos-fool-age-verification-cameras/. [Accessed: 30-May-2017].

[41] T. de Freitas Pereira *et al.*, 'Face liveness detection using dynamic texture', *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 2, Dec. 2014.

[42] L. Ghiani *et al.*, 'LivDet 2013 Fingerprint Liveness Detection Competition 2013'.

[43] J. Galbally, S. Marcel, and J. Fierrez, 'Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition', *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

[44] Z. Akhtar, C. Micheloni, and G. L. Foresti, 'Biometric Liveness Detection: Challenges and Research Opportunities', *IEEE Security & Privacy*, vol. 13, no. 5, pp. 63–72, Sep. 2015.

[45] Y. Xin *et al.*, 'A survey of liveness detection methods for face biometric systems', *Sensor Review*, pp. 00–00, May 2017.

[46] W. Kim, 'Towards real biometrics : An overview of fingerprint liveness detection', in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2016, pp. 1–3.

[47] L. Ghiani, D. A. Yambay, V. Mura, G. L. Marcialis, F. Roli, and S. A. Schuckers, 'Review of the Fingerprint Liveness Detection (LivDet) competition series: 2009 to 2015', *Image and Vision Computing*, vol. 58, pp. 110–128, 2017.

[48] D. Gragnaniello, C. Sansone, and L. Verdoliva, 'Iris liveness detection for mobile devices based on local descriptors', 2015.

[49] K. Nandakumar and A. K. Jain, 'Multibiometric Template Security Using Fuzzy Vault', in *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–6.

[50] W. Xu, Q. He, Y. Li, and T. Li, 'Cancelable voiceprint templates based on knowledge signatures', *Proceedings of the International Symposium on Electronic Commerce and Security, ISECS 2008*, no. 1, pp. 412–415, 2008.

[51] Z. Jin, A. Beng, J. Teoh, T. S. Ong, and C. Tee, 'A Revocable Fingerprint Template for Security and Privacy Preserving', *KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS*, vol. 4, no. 6, 2010.

[52] Huijuan Yang, X. Jiang, and A. C. Kot, 'Generating secure cancelable fingerprint templates using local and global features', in *2009 2nd IEEE International Conference on Computer Science and Information Technology*, 2009, pp. 645–649.

[53] S. Chikkerur, 'Generating Registration-free Cancelable Fingerprint Templates E--'.

[54] A. Juels and M. Sudan, 'A fuzzy vault scheme', in *Proceedings IEEE International Symposium on Information Theory,* p. 408.

[55] Y. Sutcu, Q. Li, and N. Memon, 'Protecting Biometric Templates With Sketch: Theory and Practice', *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 503–512, Sep. 2007.

[56] M. C. D. C. Abreu and M. Fairhurst, 'Enhancing Identity Prediction Using a Novel Approach to Combining Hard- and Soft-Biometric Information', *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 5, pp. 599–607, 2011.

[57] P. R. Anand, G. Bajpai, and V. Bhaskar, '3D Signature for Efficient Authentication in Multimodal Biometric Security Systems', *Ijetch.Org*, vol. 2, no. 2, p. 177, 2010.

[58] N. K. Ratha, J. H. Connell, and R. M. Bolle, 'Enhancing security and privacy in biometrics-based authentication systems', *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[59] S. Kanade, D. Petrovska-Delacrétaz, and B. Dorizzi, 'Cancelable iris biometrics and using error correcting codes to reduce variability in biometric data', *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pp. 120–127, 2009.

[60] O. Ouda, N. Tsumura, and T. Nakaguchi, 'Tokenless cancelable biometrics scheme for protecting iriscodes', *Proceedings - International Conference on Pattern Recognition*, no. 1, pp. 882–885, 2010.

[61] E. Maiorana, P. Campisi, J. Fierrez, J. Ortega-Garcia, and A. Neri, 'Cancelable templates for sequence-based biometrics with application to on-line signature recognition', *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 40, no. 3, pp. 525–538, 2010.

[62] J. Bringer, S. Sagem, B. Kindarji, and S. Sagem, 'Anonymous Identification with Cancelable Biometrics.pdf', pp. 494–499, 2009.

[63] J. K. J. Pillai, V. M. V. Patel, R. Chellappa, and N. K. Ratha, 'Sectored random projections for cancelable iris biometrics', *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 1838–1841, 2010.

[64] L. Menaria and K. Jain, 'A Survey on Biometric Template Protection', *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 2, no. 2, pp. 995–999, 2017.

[65] C. Rathgeb and A. Uhl, 'A survey on biometric cryptosystems and cancelable biometrics', *EURASIP Journal on Information Security*, vol. 2011, no. 1, p. 3, Dec. 2011.

[66] U. Uludag, S. Pankanti, and S. Prabhakar, 'Biometric cryptosystems: issues and challenges', *Proceedings of the*, 2004.

[67] S. Draper, A. Khisti, and E. Martinian, 'Using distributed source coding to secure fingerprint biometrics', *, Speech and Signal …*, 2007.

[68] P. Tuyls, A. Akkermans, and T. Kevenaar, 'Practical biometric authentication with template protection', *Conference on Audio- …*, 2005.

[69] Y. Dodis, L. Reyzin, and A. Smith, 'Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data', Springer, Berlin, Heidelberg, 2004, pp. 523–540.

[70] A. Juels and M. Wattenberg, 'A fuzzy commitment scheme', in *Proceedings of the 6th ACM conference on Computer and communications security - CCS '99*, 1999, pp. 28–36.

[71] T. E. Boult, W. J. Scheirer, and R. Woodwork, 'Revocable fingerprint biotokens: Accuracy and security analysis', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007.

[72] J. Z. J. Zuo, N. K. Ratha, and J. H. Connell, 'Cancelable iris biometric', *2008 19th International Conference on Pattern Recognition*, pp. 0–3, 2008.

[73] L. Leng and J. Zhang, 'PalmHash Code vs. PalmPhasor Code', *Neurocomputing*, vol. 108, pp. 1–12, 2013.

[74] M. Savvides and B. Kumar, 'Cancelable biometric filters for face recognition', *Pattern Recognition, 2004.*, 2004.

[75] J. Hämmerle-Uhl, E. Pschernig, and A. Uhl, ‘Cancelable iris biometrics using block re-mapping and image warping’, *International Conference on*, 2009.
[76] R. Bolle, J. Connell, and N. Ratha, ‘Biometric perils and patches’, *Pattern Recognition*, 2002.
[77] B. Yang, D. Hartung, and K. Simoens, ‘Dynamic random projection for biometric template protection’, in *Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on Biometrics*, 2010.
[78] Y. Kim and K. A. Toh, ‘Sparse random projection for efficient cancelable face feature extraction’, *2008 3rd IEEE Conference on Industrial Electronics and Applications, ICIEA 2008*, pp. 2139–2144, 2008.
[79] A. T. B. Jin, D. N. C. Ling, and A. Goh, ‘Biohashing: two factor authentication featuring fingerprint data and tokenised random number’, *Pattern Recognition*, vol. 37, no. 11, pp. 2245–2255, 2004.
[80] A. Teoh, A. Goh, and D. Ngo, ‘Random multispace quantization as an analytic mechanism for biohashing of biometric and random identity inputs’, *IEEE Transactions on Pattern*, 2006.
[81] T. Connie, A. Teoh, M. Goh, and D. Ngo, ‘Palmhashing: a novel approach for cancelable biometrics’, *Information processing letters*, 2005.
[82] P. Das, K. Karthik, and B. Garai, ‘A robust alignment-free fingerprint hashing algorithm based on minimum distance graphs’, *Pattern Recognition*, 2012.
[83] S. Wang and J. Hu, ‘Design of alignment-free cancelable fingerprint templates via curtailed circular convolution’, *Pattern Recognition*, vol. 47, no. 3, pp. 1321–1329, Mar. 2014.
[84] C. Karabat and H. Erdogan, ‘A cancelable biometric hashing for secure biometric verification system’, *IIH-MSP 2009 - 2009 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 1082–1085, 2009.
[85] A. M. Wing, ‘Motor control: Mechanisms of motor equivalence in handwriting’, *Current Biology*, vol. 10, no. 6, pp. R245–R248, 2000.
[86] R. Plamondon and F. J. Maarse, ‘An evaluation of motor models of handwriting’, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 5, pp. 1060–1072, 1989.
[87] D. Doermann, ‘Document image understanding: integrating recovery and interpretation’, 1993.
[88] M. Ferrer, M. Diaz, and C. Carmona-Duarte, ‘A behavioral handwriting model for static and dynamic signature synthesis’, *IEEE transactions on*, 2017.
[89] S. J. Portier, G. P. van Galen, and R. G. J. Meulenbroek, ‘Practice and the Dynamics of Handwriting Performance’, *Journal of Motor Behavior*, vol. 22, no. 4, pp. 474–492, Dec. 1990.
[90] A. Marcelli, A. Parziale, and A. Santoro, ‘Modelling Visual Appearance of Handwriting’, *International Conference on Image*, pp. 673–682, 2013.
[91] K. Franke, O. Bünnemeyer, and T. Sy, ‘Writer identification using ink texure analysis’, *Proc. 8th International Workshop on Frontiers in*, 2002.
[92] O. Hilton, ‘Effects of writing instruments on handwriting details’, *Journal of Forensic Science*, 1984.
[93] A. Thomassen and G. Van Galen, *Temporal features of handwriting: Challenges for forensic analysis*. 1996.

[94] C. Van Den Heuvel, G. van Galen, and H. Teulings, ‘Axial pen force increases with processing demands in handwriting’, *Acta psychologica*, 1998.
[95] R. van Doorn and P. Keuss, ‘Does the production of letter strokes in handwriting benefit from vision?’, *Acta Psychologica*, 1993.
[96] J. Walton, ‘Handwriting changes due to aging and Parkinson’s syndrome’, *Forensic Science International*, vol. 88, no. 3, pp. 197–214, 1997.
[97] W. G. K. Cobbah and M. Fairhurst, ‘Computer analysis of handwriting dynamics during dopamimetic tests in Parkinson’s disease’, in *Conference Proceedings of the EUROMICRO*, 2000, vol. 2, pp. 414–418.
[98] M. Uzun, N. Alkan, O. Kurtas, R. Yilmaz, M. Can, and I. Birincioglu, ‘Changes in Handwriting due to Alzheimers Disease a Case Report’, *The Bulletin of Legal Medicine*, vol. 21, no. 2, Aug. .
[99] B. C. M. Smits-Engelsman and G. P. Van Galen, ‘Dysgraphia in Children: Lasting Psychomotor Deficiency or Transient Developmental Delay?’, *Journal of Experimental Child Psychology*, vol. 67, no. 2, pp. 164–184, 1997.
[100] J. Neils-Strunjas, K. Groves-Wright, P. Mashima, S. Harnish, W. T. F., and W. G. J., ‘Dysgraphia in Alzheimer’s Disease: A Review for Clinical and Research Purposes’, *Journal of Speech Language and Hearing Research*, vol. 49, no. 6, p. 1313, Dec. 2006.
[101] L. Schomaker, ‘Advances in Writer Identification and Verification’, in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2*, 2007, pp. 1268–1273.
[102] L. Schomaker and M. Bulacu, ‘Automatic writer identification using connected-component contours and edge-based features of uppercase western script’, *IEEE Transactions on Pattern*, 2004.
[103] M. Simner, A. Marcelli, S. Ablameyko, and K. Lange, ‘A Comparison of Arabic Numerical Allographs Written by Adults from English Speaking vs. Non-Engilish Speaking Countries’, *Proceedings of the 11th*, 2003.
[104] R. Sassoon, *Handwriting : the way to teach it*. Paul Chapman Pub., 2003.
[105] Saad Mamoun Abdel Rhman Ahmed, ‘Off-Line Arabic Signature Verification Using Geometrical Features’, in *National Workshop on Information Assurance Research (WIAR ’2012).*, 2012.
[106] A. Soleimani, K. Fouladi, and B. N. Araabi, ‘Persian offline signature verification based on curvature and gradient histograms’, in *2016 6th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2016, pp. 147–152.
[107] S. Pal, V. Nguyen, M. Blumenstein, and U. Pal, ‘Off-Line Bangla Signature Verification’, in *2012 10th IAPR International Workshop on Document Analysis Systems*, 2012, pp. 282–286.
[108] H.-D. Chang, J.-F. Wang, and H.-M. Suen, ‘Dynamic handwritten Chinese signature verification’, *Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR ’93)*, no. MDi, pp. 2–5, 1993.
[109] H. Lv, W. Wang, C. Wang, and Q. Zhuo, ‘Off-line Chinese signature verification based on support vector machines’, 2005.
[110] S. Pal, A. Alireza, U. Pal, and M. Blumenstein, ‘Multi-script off-line signature identification’, in *2012 12th International Conference on Hybrid Intelligent Systems (HIS)*, 2012, pp. 236–240.

[111] S. Pal, U. Pal, and M. Blumenstein, ‘Off-line English and Chinese signature identification using foreground and background features’, in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012, pp. 1–7.
[112] D.-Y. Yeung *et al.*, ‘SVC2004: First International Signature Verification Competition’.
[113] M. Liwicki *et al.*, ‘Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011)’, in *2011 International Conference on Document Analysis and Recognition*, 2011, pp. 1480–1484.
[114] A. K. Jain, S. C. Dass, and K. Nandakumar, ‘Can soft biometric traits assist user recognition?’, 2004, pp. 561–572.
[115] M. Abreu and M. Fairhurst, ‘Improving Identity Prediction in Signature-based Unimodal Systems Using Soft Biometrics’, Springer, Berlin, Heidelberg, 2009, pp. 348–356.
[116] M. Erbilek, ‘Improved age prediction from biometric data using multimodal configurations’, pp. 179–186.
[117] N. Bouadjenek, H. Nemmour, and Y. Chibani, ‘Robust soft-biometrics prediction from off-line handwriting analysis’, *Applied Soft Computing*, vol. 46, pp. 980–990, 2016.
[118] I. Yoshimura and M. Yoshimura, ‘EVALUATION OF SIGNATURE QUALITY AS A FUNCTION OF NATIONALITY VIA AN OFF-LINE SIGNATURE VERIFICATION SYSTEM’, *Intelligent Automation and Soft Computing*, vol. 7, no. 3, pp. 195–203, 2001.
[119] D. Impedovo and G. Pirlo, ‘Automatic Signature Verification: The State of the Art’, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 5, pp. 609–635, Sep. 2008.
[120] D. Hamilton, J. Whelan, A. McLaren, and I. MacIntyre, ‘Low cost dynamic signature verification system’, 1995.
[121] F. R. Rioja, M. N. Miyatake, H. P. Meana, and K. Toscano, ‘Dynamic features extraction for on-line signature verification’, *14th International Conference on Electronics, Communications and Computers, 2004. CONIELECOMP 2004.*, 2004.
[122] F. Alonso-Fernandez and J. Fierrez-Aguilar, ‘On-line signature verification using Tablet PC’, *Image and Signal*, 2005.
[123] F. Alonso-Fernandez, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, ‘Secure access system using signature verification over tablet PC’, *IEEE Aerospace and Electronic Systems Magazine*, vol. 22, no. 4, pp. 3–8, Apr. 2007.
[124] J. Ortega-Garcia *et al.*, ‘MCYT baseline corpus: a bimodal biometric database’, *IEE Proceedings - Vision, Image, and Signal Processing*, vol. 150, no. 6, p. 395, 2003.
[125] R. Blanco-Gonzalo, O. Miguel-Hurtado, A. Mendaza-Ormaza, and R. Sanchez-Reillo, ‘Handwritten signature recognition in mobile scenarios: Performance evaluation’, in *2012 IEEE International Carnahan Conference on Security Technology (ICCST)*, 2012, pp. 174–179.
[126] M. Djioua and R. Plamondon, ‘Studying the variability of handwriting patterns using the Kinematic Theory’, *Human Movement Science*, vol. 28, no. 5, pp. 588–601, 2009.
[127] T. Qu, A. El Saddik, and A. Adler, ‘Dynamic signature verification system using stroked based features’, *Haptic, Audio and Visual*, 2003.

[128] Napa Sae-Bae and N. Memon, 'Online Signature Verification on Mobile Devices', *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, pp. 933–947, Jun. 2014.
[129] S. Shirato and D. Muramatsu, 'Camera-based online signature verification system: effects of camera positions', *Congress (WAC), 2010*, 2010.
[130] D. Muramatsu and K. Yasuda, 'Biometric person authentication method using camera-based online signature acquisition', in *Document Analysis and*, 2009.
[131] Y. Fang, W. Kang, Q. Wu, and L. Tang, 'A novel video-based system for in-air signature verification', *Computers & Electrical Engineering*, vol. 57, pp. 1–14, 2017.
[132] M. I. Malik, M. Liwicki, and A. Dengel, 'Evaluation of Local and Global Features for Offline Signature Verification', in *1st International Workshop on Automated Forensic Handwriting Analysis (AFHA)*, 2011, pp. 26–30.
[133] H. Baltzakis and N. Papamarkos, 'A new signature verification technique based on a two-stage neural network classifier', *Engineering applications of Artificial*, 2001.
[134] G. Dimauro, S. Impedovo, and G. Pirlo, 'A multi-expert signature verification system for bankcheck processing', *International Journal of*, 1997.
[135] O. Yedekcoglu, M. Akban, and Y. Lim, 'Off-line signature verification with thickened templates', *COMCON 5} Proceedings of 5th International*, 1995.
[136] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, 'Offline Handwritten Signature Verification - Literature Review'.
[137] M. Zou, J. Tong, C. Liu, and Z. Lou, 'On-line signature verification using local shape analysis', *Document Analysis and*, 2003.
[138] S. Chi, J. Lee, J. Soh, D. Kim, W. Oh, and C. Kim, 'Visualization of Dynamic Characteristics in Two-Dimensional Time Series Patterns: An Application to Online Signature Verification', Springer, Berlin, Heidelberg, 2004, pp. 395–409.
[139] G. Congedo, G. Dimauro, S. Impedovo, and G. Pirlo, 'Off-line signature verification by fundamental components analysis', 1993.
[140] M. Ammar, Y. Yoshida, and T. Fukumura, 'Structural description and classification of signature images', *Pattern recognition*, 1990.
[141] R. Sabourin and R. Plamondon, 'Segmentation of handwritten signature images using the statistics of directional data', *Pattern Recognition, 1988., 9th*, 1988.
[142] R. PLAMONDON, 'The design of an on-line signature verification sytem: from theory to practice', *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 8, no. 3, pp. 795–811, Jun. 1994.
[143] H. Y. Kwon, E. Y. Ha, and H. Y. Hwang, 'Online Signature Verification Based on Dynamic Feature Segmentation and 3-Step Matching', Springer, Berlin, Heidelberg, 2003, pp. 1062–1065.
[144] J. Lee, H. Yoon, J. Soh, B. Chun, and Y. Chung, 'Using geometric extrema for segment-to-segment characteristics comparison in online signature verification', *Pattern Recognition*, 2004.
[145] 'Optical scanner'. [Online]. Available: http://uk.pcmag.com/epson-perfection-v850-pro/39018/review/epson-perfection-v850-pro. [Accessed: 25-Jan-2018].
[146] 'WACOM Intuos3 User's Manual for Windows & Macintosh'. [Online]. Available: https://www.uwo.ca/visarts/resources/Sign-Out-Guides/Digital Drawing Board/Intuos3.pdf. [Accessed: 19-May-2016].

[147] A. Fischer, M. Diaz, R. Plamondon, and M. A. Ferrer, 'Robust score normalization for DTW-based on-line signature verification', in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015, pp. 241–245.

[148] J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez, 'HMM-based on-line signature verification: Feature extraction and signature modeling', *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2325–2334, 2007.

[149] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Peñalba, J. Ortega-Garcia, and D. Maltoni, 'An On-Line Signature Verification System Based on Fusion of Local and Global Information', Springer, Berlin, Heidelberg, 2005, pp. 523–532.

[150] G. Congedo, G. Dimauro, A. Forte, and S. Impedovo, 'Selecting reference signatures for on-line signature verification', *Conference on Image …*, 1995.

[151] Zhili Chen, Xinghua Xia, and Fangjun Luan, 'Automatic online signature verification based on dynamic function features', in *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2016, pp. 964–968.

[152] T. Islam and M. Fairhurst, 'Natural revocability in handwritten signatures to enhance biometric security', in *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*, 2012, pp. 791–796.

[153] M. Erbilek and M. Fairhurst, 'Framework for managing ageing effects in signature biometrics', *IET Biometrics*, vol. 1, no. 2, p. 136, 2012.

[154] C. Rabasse, R. M. R. M. Guest, and M. C. Fairhurst, 'Data With Natural Variability', *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 3, pp. 691–699, Jun. 2008.

[155] M. Razian, M. Fairhurst, and S. Hoque, 'Effect of dynamic features on diagnostic testing for dyspraxia', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3118, pp. 1039–1046, 2004.

[156] B. Yanikoglu and A. Kholmatov, 'Online signature verification using fourier descriptors', *Eurasip Journal on Advances in Signal Processing*, vol. 2009, 2009.

[157] S. Wang, G. Deng, and J. Hu, 'A partial Hadamard transform approach to the design of cancelable fingerprint templates containing binary biometric representations', *Pattern Recognition*, vol. 61, pp. 447–458, 2017.

[158] X. Song, X. Xia, and F. Luan, 'Online Signature Verification Based on Stable Features Extracted Dynamically', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–14, 2016.

[159] Y. Liu, Z. Yang, and L. Yang, 'Online signature verification based on DCT and sparse representation', *IEEE transactions on cybernetics*, 2015.

[160] E. N. Zois, L. Alewijnse, and G. Economou, 'Offline signature verification and quality characterization using poset-oriented grid features', *Pattern Recognition*, vol. 54, pp. 162–177, Jun. 2016.

[161] K. E. Mwangi, 'Offline Handwritten Signature Verification Using Sift Features', 2008.

[162] M. Fairhurst and P. Brittan, 'An evaluation of parallel strategies for feature vector construction in automatic signature verification systems', *journal of pattern recognition and artificial …*, 1994.

[163] A. Marcano-Cedeno, J. Quintanilla-Dominguez, M. G. Cortina-Januchs, and D. Andina, 'Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network', in *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, 2010, pp. 2845–2850.

[164] F. Song, Z. Guo, and D. Mei, 'Feature Selection Using Principal Component Analysis', in *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, 2010, pp. 27–30.

[165] M. Michael and W.-C. Lin, 'Experimental Study of Information Measure and Inter-Intra Class Distance Ratios on Feature Selection and Orderings', *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 2, pp. 172–181, 1973.

[166] A. Verikas and M. Bacauskiene, 'Feature selection with neural networks', 2002.

[167] J. Livarinen, K. Valkealahti, A. Visa, and O. Simula, 'Feature Selection with Self-Organizing Feature Map', *International Conference on Artificial*, 1994.

[168] J. Yang and V. Honavar, 'Feature subset selection using a genetic algorithm', *IEEE Intelligent Systems*, vol. 13, no. 2, pp. 44–49, Mar. 1998.

[169] A. G. Karegowda, a S. Manjunath, M. a Jayaram, A. Gowda Karegowda, a S. Manjunath, and M. a Jayaram, 'Comparative study of attribute selection using gain ration and correlation based feature selection', *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 271–277, 2010.

[170] M. Parizeau and R. Plamondon, *Comparative analysis of regional correlation, dynamic time warping, and skeletal tree matching for signature verification*, vol. 12, no. 7. 1990, pp. 710–716.

[171] Y. Chen and X. Ding, 'On-line signature verification using direction sequence string matching', *International Conference on …*, 2002.

[172] Y. Mizukami, M. Yoshimura, and H. Miike, 'An off-line signature verification system using an extracted displacement function', *Pattern Recognition*, 2002.

[173] K. Huang and H. Yan, 'Off-line signature verification using structural feature correspondence', *Pattern Recognition*, 2002.

[174] B. C. Madasu, Vamsi Krishna and Lovell, 'An Automatic Off-Line Signature Verification and Forgery Detection System', *Pattern Recognition Technologies and Applications: Recent Advances*, pp. 63–89, 2008.

[175] B. Kar and P. Dutta, 'SVM based signature verification by fusing global and functional features', *International Journal of …*, vol. 60, no. 16, pp. 34–39, 2012.

[176] Siyuan Chen and S. Srihari, 'Use of exterior contours and shape features in off-line signature verification', in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, 2005, p. 1280–1284 Vol. 2.

[177] J. Galbally, J. Fierrez, and J. Ortega-Garcia, 'Classification of handwritten signatures based on name legibility - art. no. 653907', *Biometric Technology for Human Identification IV*, vol. 6539, p. 53907, 2007.

[178] R. Sabourin and J. Drouhard, 'Off-line signature verification using directional PDF and neural networks', *1992. Vol. II. Conference B: Pattern …*, 1992.

[179] K. Huang and H. Yan, 'Stability and style-variation modeling for on-line signature verification', *Pattern Recognition*, vol. 36, no. 10, pp. 2253–2270, 2003.

[180] F. Roli, 'Multiple Classifier Systems', in *Encyclopedia of Biometrics*, Boston, MA: Springer US, 2015, pp. 1142–1147.

[181] L. Batista, E. Granger, and R. Sabourin, 'A Multi-Classifier System for Off-Line Signature Verification Based on Dissimilarity Representation', Springer, Berlin, Heidelberg, 2010, pp. 264–273.

[182] G. S. Eskander, E. Granger, and R. Sabourin, ‘Hybrid writer-independent–writer-dependent offline signature verification system’, *IET Biometrics*, vol. 2, no. 4, pp. 169–181, Dec. 2013.

[183] C. O’Reilly and R. Plamondon, ‘Linking brain stroke risk factors to human movement features for the development of preventive tools’, *Frontiers in aging*, 2014.

[184] R. Plamondon, C. O’Reilly, and C. Ouellet-Plamondon, ‘Strokes against stroke—strokes for strides’, *Pattern Recognition*, 2014.

[185] A. Ginestroni, S. Diciotti, P. Cecchi, and I. Pesaresi, ‘Neurodegeneration in friedreich’s ataxia is associated with a mixed activation pattern of the brain. A fMRI study’, *Human brain*, 2012.

[186] U. Rüb, D. Del Turco, K. Del Tredici, and R. De Vos, ‘Thalamic involvement in a spinocerebellar ataxia type 2 (SCA2) and a spinocerebellar ataxia type 3 (SCA3) patient, and its clinical relevance’, *Brain*, 2003.

[187] C. Simonet, A. J. Noyce, H. Ling, A. J. Lees, and T. T. Warner, ‘Handwriting in Parkinson’’, *Movement Disorders*, vol. 31, p. S508, 2016.

[188] J. G. Phillips, J. L. Bradshaw, E. Chiu, and J. A. Bradshaw, ‘Characteristics of handwriting of patients with huntington’s disease’, *Movement Disorders*, vol. 9, no. 5, pp. 521–530, 1994.

[189] Y. Zhang, B. Tang, A. Zhao, K. Xia, and Z. Long, ‘Novel compound heterozygous mutations in the PANK2 gene in a Chinese patient with atypical pantothenate kinase-associated neurodegeneration’, *Movement*, 2005.

[190] B. Ghali, N. Thalanki Anantha, J. Chan, T. Chau, and G. Schöner, ‘Variability of Grip Kinetics during Adult Signature Writing’, *PLoS ONE*, vol. 8, no. 5, p. e63216, May 2013.

[191] W. Guerfali and R. Plamondon, ‘A new method for the analysis of simple and complex planar rapid movements’, *Journal of Neuroscience Methods*, vol. 82, pp. 35–45, 1998.

[192] R. Plamondon and W. Guerfali, ‘The generation of handwriting with delta-lognormal synergies’, *Biological Cybernetics*, vol. 78, no. 2, pp. 119–132, Feb. 1998.

[193] C. O’Reilly and R. Plamondon, ‘Impact of the principal stroke risk factors on human movements’, *Human movement science*, 2011.

[194] H. Ploog, *Handwriting psychology : personality reflected in handwriting*. iUniverse Com, 2013.

[195] K. Anderson and P. W. McOwan, ‘Real-Time Emotion Recognition Using Biologically Inspired Models’, 2003, pp. 119–127.

[196] M. Fairhurst, M. Erbilek, C. Li, and D. Arts, ‘ENHANCING THE FORENSIC VALUE OF HANDWRITING USING EMOTION’, 2014.

[197] O. Hilton, *Scientific examination of questioned documents*. CRC press, 1993.

[198] R. Plamondon and G. Lorette, ‘Automatic signature verification and writer identification — the state of the art’, *Pattern Recognition*, vol. 22, no. 2, pp. 107–131, Jan. 1989.

[199] R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, A. Acien, and R. Tolosana, ‘e-BioSign tool: Towards scientific assessment of dynamic signatures under forensic conditions’, in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2015, pp. 1–6.

[200] K. Franke, L. Schomaker, C. Veenhuis, L. Vuurpijl, M. van Erp, and I. Guyon, 'WANDA: A common ground for forensic handwriting examination and writer identification', *ENFHEX news-Bulletin of the European Network of Forensic Handwriting Experts*, vol. 1, no. 4, pp. 23–47, 2004.
[201] M. C. D. Abreu and M. C. Fairhurst, 'Improving Forgery Detection in Off-Line Forensic Signature Processing', 2009.
[202] S. N. Srihari, S.-H. Cha, H. Arora, and S. Lee, 'Individuality of Handwriting', *Journal of Forensic Sciences*, vol. 47, no. 4, p. 15447J, Jul. 2002.
[203] T. Islam and M. Fairhurst, 'Investigating an objective measure of writer hesitation for forensic analysis of the handwritten signature', in *2016 4th International Conference on Biometrics and Forensics (IWBF)*, 2016, pp. 1–6.
[204] A. Jain, R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society*. Springer Science \& Business Media, 2006.
[205] D. Impedovo, G. Pirlo, and R. Plamondon, 'Handwritten signature verification: New advancements and open issues', *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*, pp. 367–372, 2012.
[206] F. Leclerc and R. Plamondon, 'Automatic Signature Verification: The State of the Art -- 1989-1993', *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 8, no. 3, pp. 643–660, 1994.
[207] R. M. Guest, 'The repeatability of signatures', *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*, pp. 492–497, 2004.
[208] 'MEDDRAW - Computer-Based Drawing Task'. [Online]. Available: https://www.eda.kent.ac.uk/research/theme_project.aspx?pid=59. [Accessed: 19-May-2016].
[209] M. Fairhurst and M. Abreu, 'An investigation of predictive profiling from handwritten signature data', *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 1305–1309, 2009.
[210] H. Lei and V. Govindaraju, 'A comparative study on the consistency of features in on-line signature verification', *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2483–2489, Nov. 2005.
[211] C. Rabasse, R. M. Guest, and M. C. Fairhurst, 'A New Method for the Synthesis of Signature Data With Natural Variability', *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 3, pp. 691–699, Jun. 2008.
[212] S. Impedovo and G. Pirlo, 'Verification of handwritten signatures: An overview', in *Proceedings - 14th International conference on Image Analysis and Processing, ICIAP 2007*, 2007, pp. 191–196.
[213] R. Guest, 'Age dependency in handwritten dynamic signature verification systems', 2006.
[214] L. Mohammed, B. Found, M. Caligiuri, and D. Rogers, 'PAPER QUESTIONED DOCUMENTS Dynamic Characteristics of Signatures: Effects of Writer Style on Genuine and Simulated Signatures*'.
[215] J. Fierrez and J. Ortega-Garcia, 'On-Line Signature Verification', in *Handbook of Biometrics*, Boston, MA: Springer US, 2008, pp. 189–209.
[216] C. Spearman, 'The Proof and Measurement of Association between Two Things', *Source: The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.
[217] P. Cunningham and S. J. Delany, 'k-Nearest Neighbour Classifiers', 2007.
[218] S. Arya, D. M. Mount, and Netanyahu, 'An optimal algorithm for approximate nearest neighbor searching fixed dimensions'.

[219] S. Arya *et al.*, 'An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions', pp. 573–582, 1994.
[220] Weka, 'Weka 3 - Data Mining with Open Source Machine Learning Software in Java'. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/. [Accessed: 06-Oct-2015].
[221] J. Richiardi, A. Drygajlo, and L. Todesco, 'Promoting diversity in Gaussian mixture ensembles: An application to signature verification', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5372 LNCS, pp. 140–149, 2008.
[222] S. Garcia-Salicetti *et al.*, 'BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities', Springer Berlin Heidelberg, 2003, pp. 845–853.
[223] D. Y. et al. Yeung, 'SVC2004: First international signature verification competition', in *Biometric Authentication*, 2004, pp. 16–22.
[224] Biosecure, 'Biometrics for Secure Authentication, FP6 NoE, IST- 2002-507634', 2004.
[225] J. Ortega-Garcia *et al.*, 'Software tool and acquisition equipment recommendations for the three scenarios considered', *Universidad Politechnica de Madrid, Tech. Rep. D*, vol. 6, 2006.
[226] M. E. Munich and P. Perona, 'Camera-Based ID Verification by Signature Tracking', pp. 782–796, Jun. 1998.
[227] M. E. Munich and P. Perona, 'Visual identification by signature tracking', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 200–217, Feb. 2003.
[228] G. Lorette and R. Plamondon, 'Dynamic approaches to handwritten signature verification', *Computer Processing of Handwriting*, pp. 21–47, 1990.
[229] M. L. Meeks and T. T. Kuklinski, 'Measurement of dynamic digitizer performance', *Computer Processing of Handwriting*, pp. 89–110, 1990.
[230] J. Fierrez *et al.*, 'BiosecurID: a multimodal biometric database', *Pattern Analysis and Applications*, vol. 13, no. 2, pp. 235–246, May 2010.
[231] 'intuos3-users-manual'. .
[232] Wacom Technology Corporation, 'Wintab Interface Specification 1.4': [Online]. Available: http://www.wacomeng.com/windows/docs/Wintab_v140.htm#_Toc275759760. [Accessed: 19-Apr-2016].
[233] F. Q. F. Quan, S. F. S. Fei, C. A. C. Anni, and Z. F. Z. Feifei, 'Cracking Cancelable Fingerprint Template of Ratha', *2008 International Symposium on Computer Science and Computational Technology*, vol. 2, pp. 572–575, 2008.
[234] G. Aggarwal, N. K. Ratha, J. H. Connell, and R. M. Bolle, 'Physics-Based Revocable Face Recognition University of Maryland , College Park IBM T . J . Watson Research Center', pp. 5232–5235, 2008.
[235] O. OUDA, N. TSUMURA, and T. NAKAGUCHI, 'BioEncoding: A Reliable Tokenless Cancelable Biometrics Scheme for Protecting IrisCodes', *IEICE TRANSACTIONS on Information and Systems*, vol. E93–D, no. 7, pp. 1878–1888, 2010.
[236] K. Yang, Y. Du, Z. Zhou, and C. Belcher, 'Gabor Descriptor based cancelable iris recognition method', in *2010 IEEE International Conference on Image Processing*, 2010, pp. 4085–4088.

[237] B. Choudhury, P. Then, V. Raman, B. Issac, and M. K. Haldar, 'Cancelable iris Biometrics based on data hiding schemes', in *2016 IEEE Student Conference on Research and Development (SCOReD)*, 2016, pp. 1–6.
[238] H. Baier, F. Breitinger, C. Busch, and C. Rathgeb, 'On application of bloom filters to iris biometrics', *IET Biometrics*, vol. 3, no. 4, pp. 207–218, Dec. 2014.
[239] M. Gomez-Barrero, C. Rathgeb, J. Galbally, C. Busch, and J. Fierrez, 'Unlinkable and irreversible biometric template protection based on bloom filters', *Information Sciences*, vol. 370371, pp. 18–32, 2016.
[240] C. Rathgeb and C. Busch, 'Cancelable multi-biometrics: Mixing iris-codes based on adaptive bloom filters', *Computers & Security*, vol. 42, pp. 1–12, 2014.
[241] V. S. Meenakshi and D. G. Padmavathi, 'Security analysis of password hardened multimodal biometric fuzzy vault with combined feature points extracted from fingerprint, iris and retina for high security applications', *Procedia Computer Science*, vol. 2, pp. 195–206, 2010.
[242] H. Yang, X. Jiang, and A. C. Kot, 'Generating Secure Cancelable Fingerprint Templates using locan global features.pdf', pp. 0–4, 2009.
[243] W. J. Wong, A. B. J. Teoh, Y. H. Kho, and M. L. D. Wong, 'Kernel PCA enabled bit-string representation for minutiae-based cancellable fingerprint template', *Pattern Recognition*, vol. 51, pp. 197–208, 2016.
[244] M. Damasceno and A. M. P. Canuto, 'An empirical analysis of ensemble systems in cancellable behavioural biometrics: A touch screen dataset', in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 2661–2668.
[245] Y. Xie, F. Zhang, X. Chen, and K. Kim, 'ID-Based Distributed "Magic Ink" Signature from Pairings', Springer Berlin Heidelberg, 2003, pp. 249–259.
[246] M. R. Freire, J. Fierrez, J. Galbally, and J. Ortega-Garcia, 'Biometric Hashing Based on Genetic Selection and Its Application to On-Line Signatures', in *Advances in Biometrics*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1134–1143.
[247] M. Fouad, F. Alsulaiman, A. El Saddik, and E. Petriu, 'Revocable handwritten signatures with haptic information', in *2011 IEEE International Workshop on Haptic Audio Visual Environments and Games*, 2011, pp. 108–111.
[248] 'Proc. of Securtech,Washington D.C., April 7-9, 1992.', in *Proc. of Securtech, Washington D.C., April 7-9, 1992.*
[249] I. W. Evett and R. N. Totty, 'A Study of the Variation in the Dimensions of Genuine Signatures', *Journal of the Forensic Science Society*, vol. 25, no. 3, pp. 207–215, May 1985.
[250] M. A. Eldridge, I. Nimmo-Smith, A. M. Wing, and R. N. Totty, 'The Variability of Selected Features in Cursive Handwriting: Categorical Measures', *Journal of the Forensic Science Society*, vol. 24, no. 3, pp. 179–219, May 1984.
[251] M. Diaz, A. Fischer, M. A. Ferrer, and R. Plamondon, 'Dynamic Signature Verification System Based on One Real Signature', *IEEE Transactions on Cybernetics*, pp. 1–12, 2016.
[252] R. Sabourin, G. Genest, and F. J. Preteux, 'Off-line signature verification by local granulometric size distributions', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 976–988, 1997.
[253] 'Automatic signatue verification:The State of the Art'. [Online]. Available: http://www.inf.ufpr.br/lesoliveira/padroes/Impedovo08.pdf. [Accessed: 06-Oct-2015].

[254] G. Doddington, W. Liggett, and A. Martin, 'Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation', ... *International Conference on ...*, pp. 1–4, 1998.
[255] S. Garcia-Salicetti *et al.*, 'Online Handwritten Signature Verification'.
[256] R. Mergl, P. Tigges, A. Schröter, H.-J. Möller, and U. Hegerl, 'Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives', *Journal of Neuroscience Methods*, vol. 90, no. 2, pp. 157–169, Aug. 1999.
[257] M. Djioua and R. Plamondon, 'An interactive system for the automatic generation of huge handwriting databases from a few specimens', in *2008 19th International Conference on Pattern Recognition*, 2008, pp. 1–4.
[258] G. Pirlo and D. Impedovo, 'Verification of Static Signatures by Optical Flow Analysis', *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 5, pp. 499–505, Sep. 2013.
[259] Z. Smekal, J. Mekyska, I. Rektorova, and M. Faundez-Zanuy, 'Analysis of neurological disorders based on digital processing of speech and handwritten text', in *International Symposium on Signals, Circuits and Systems ISSCS2013*, 2013, pp. 1–6.
[260] M. Gomez-Barrero, J. Galbally, J. Fierrez, J. Ortega-Garcia, and R. Plamondon, 'Variations of handwritten signatures with time: A sigma-lognormal analysis', *Proceedings - 2013 International Conference on Biometrics, ICB 2013*, 2013.
[261] G. Pirlo and D. Impedovo, 'Cosine similarity for analysis and verification of static signatures', *IET Biometrics*, vol. 2, no. 4, pp. 151–158, 2013.
[262] D. Impedovo, G. Pirlo, L. Sarcinella, E. Stasolla, and C. A. Trullo, 'Analysis of Stability in Static Signatures Using Cosine Similarity', in *2012 International Conference on Frontiers in Handwriting Recognition*, 2012, pp. 231–235.
[263] G. Pirlo, D. Impedovo, E. Stasolla, and C. A. Trullo, 'Learning Local Correspondences for Static Signature Verification', Springer Berlin Heidelberg, 2009, pp. 385–394.
[264] S. Müller and O. Henniger, 'Evaluating the Biometric Sample Quality of Handwritten Signatures', in *Advances in Biometrics*, vol. 4642, S.-W. Lee and S. Z. Li, Eds. Springer Berlin Heidelberg, 2007, pp. 407–414.
[265] N. Yager and T. Dunstone, 'The Biometric Menagerie', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 220–230, Feb. 2010.
[266] N. Yager and T. Dunstone, 'Worms, Chameleons, Phantoms and Doves: New Additions to the Biometric Menagerie', in *2007 IEEE Workshop on Automatic Identification Advanced Technologies*, 2007, pp. 1–6.
[267] N. Poh and J. Kittler, 'A methodology for separating sheep from goats for controlled enrollment and multimodal fusion', in *2008 Biometrics Symposium*, 2008, pp. 17–22.
[268] M. Wittman, P. Davis, and P. J. Flynn, 'Empirical Studies of the Existence of the Biometric Menagerie in the FRGC 2.0 Color Image Corpus', in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pp. 33–33.
[269] S. Mumtazah *et al.*, 'ANALYSIS OF GOAT WITHIN USER POPULATION OF AN OFFLINE SIGNATURE BIOMETRICS Intra - user variability inherent in human handwritten signatures remains one of the main challenges for a robust biometrics signature based authentication system . The existe', no. Isspa, pp. 765–769, 2010.

[270] N. Poh, J. Kittler, A. Rattani, and M. Tistarelli, 'Group-specific score normalization for biometric systems', in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010, pp. 38–45.

[271] J. Paone and P. J. Flynn, 'On the consistency of the biometric menagerie for irises and iris matchers', in *2011 IEEE International Workshop on Information Forensics and Security*, 2011, pp. 1–6.

[272] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines : and other kernel-based learning methods*. Cambridge University Press, 2000.

[273] Chih-Wei Hsu and Chih-Jen Lin, 'A comparison of methods for multiclass support vector machines', *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[274] R. Kohavi, 'A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection', 1995.

[275] L. L. Lee, T. Berger, and E. Aviczer, 'Reliable online human signature verification systems', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 643–647, Jun. 1996.

[276] K. R. Radhika, G. N. Sekhar, and M. K. Venkatesha, 'Pattern recognition techniques in on-line hand written signature verification - A survey', in *2009 International Conference on Multimedia Computing and Systems*, 2009, pp. 216–221.

[277] R. Plamondon, L. De Yu, G. E. Stelmach, and B. Clement, 'On the automatic extraction of biomechanical information from handwriting signals', *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 1, pp. 90–101, 1991.

[278] M. Parodi and J. C. Gómez, 'Legendre polynomials based feature extraction for online signature verification. Consistency analysis of feature combinations', *Pattern Recognition*, vol. 47, no. 1, pp. 128–140, 2013.

[279] J. Galbally, J. Fierrez, M. Martinez-Diaz, and R. Plamondon, 'Quality Analysis of Dynamic Signature Based on the Sigma-Lognormal Model', in *2011 International Conference on Document Analysis and Recognition*, 2011, pp. 633–637.

[280] L. R. B. Schomaker and R. Plamondon, 'The relation between pen force and pen-point kinematics in handwriting', *Biological Cybernetics*, vol. 63, no. 4, pp. 277–289, Aug. 1990.

[281] C. Vielhauer and R. Steinmetz, 'Handwriting: Feature correlation analysis for biometric hashes', *Eurasip Journal on Applied Signal Processing*, vol. 2004, no. 4, pp. 542–558, 2004.

[282] L. A. Mohammed, B. Found, M. Caligiuri, and D. Rogers, 'The dynamic character of disguise behavior for text-based, mixed, and stylized signatures.', *Journal of forensic sciences*, vol. 56 Suppl 1, pp. S136-41, Jan. 2011.

[283] F. Alonso-Fernandez and M. Fairhurst, 'Impact of signature legibility and signature type in off-line signature verification', *Biometrics*, 2007.

[284] O. Miguel-Hurtado, R. Guest, S. V. Stevenage, and G. J. Neil, 'The relationship between handwritten signature production and personality traits', in *IEEE International Joint Conference on Biometrics*, 2014, pp. 1–8.

[285] J. Robertson and R. Guest, 'A feature based comparison of pen and swipe based signature characteristics', *Human Movement Science*, vol. 43, pp. 169–182, 2015.

[286] N. Houmani and S. Garcia-Salicetti, 'Quality Measures for Online Handwritten Signatures', 2014, pp. 255–283.

[287] N. E. Fenton and M. (Martin D. . Neil, *Risk assessment and decision analysis with Bayesian networks*. CRC press, 2012.

[288] D. E. Hinkle, W. Wiersma, and S. G. Jurs, *Applied statistics for the behavioral sciences*. Wadsworth, Cengage Learning, 2009.
[289] C. E. Brown, 'Coefficient of Variation', in *Applied Multivariate Statistics in Geohydrology and Related Sciences*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 155–157.
[290] H. Abdi and L. J. Williams, 'Tukey's Honestly Significant Difference (HSD) Test', 2010.
[291] 'Questioned documents by Osborn'. [Online]. Available: http://www.jstor.org/stable/1147312?seq=1#page_scan_tab_contents. [Accessed: 06-Oct-2015].
[292] T. Ostrum, B and Tanaka, 'Another look at handwriting movement', *Journal of the American Society of Questioned Document Examiners*, vol. 9, no. 2, pp. 57–67, 2006.
[293] T. Guest, Richard; Fairhurst, M; Linnell, 'Towards an Inferred Data Accuracy Assessment of Forensic Document Examination Methodologies for Signatures', *Proceedings of the International Graphonomics Society*, 2009.
[294] A. Van Galen, GP and Van Gemmert, 'Kinematic and dynamic features of forging another person's handwriting', *Journal of Forensic Document Examination*, vol. 9, pp. 1–25, 1996.
[295] M. G. Longstaff and R. A. Heath, 'A nonlinear analysis of the temporal characteristics of handwriting', *Human Movement Science*, vol. 18, no. 4, pp. 485–524, 1999.
[296] L. Harralson, Heidi H and Miller, 'Developments in Handwriting and Signature Identification in the Digital Age', 2014. [Online]. Available: https://scholar.google.co.uk/scholar.bib?q=info:-jeqMJ2VpWgJ:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAVhPENUNH-6Zry5EvqxyszsWdK9iUa7k5&scisf=4&hl=en&scfhb=1. [Accessed: 06-Oct-2015].
[297] I. ASTM., 'E2290-07a Standard Guide for Examination of Handwritten Items', 2007. [Online]. Available: http://www.fbi.gov/cgi-bin/outside.cgi?http://www.astm.org/Standards/E2290.htm. [Accessed: 06-Oct-2015].
[298] K. Franke, 'Analysis of authentic signatures and forgeries', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5718 LNCS, pp. 150–164, 2009.
[299] G. Brand, Max and Brand, *Practical fluency: Classroom perspectives, grades k-6*. Stenhouse Publishers, 2006.
[300] A. MacGowan-Gilhooly, 'Fluency first: Reversing the traditional ESL sequence', *Journal of Basic Writing*, vol. 10, no. 1, pp. 73–87, 1991.
[301] C. P. Casanave, *Controversies in second language writing: Dilemmas and decisions in research and instruction*. University of Michigan Press, 2004.
[302] A. Overvelde and W. Hulstijn, 'Handwriting development in grade 2 and grade 3 primary school children with normal, at risk, or dysgraphic characteristics.', *Research in developmental disabilities*, vol. 32, no. 2, pp. 540–8, Jan. 2011.
[303] D. McCUTCHEN, '"Functional Automaticity" in Children's Writing: A Problem of Metacognitive Control', *Written Communication*, vol. 5, no. 3, pp. 306–324, Jul. 1988.

[304] R. G. J. Meulenbroek and A. W. A. Van Gemmert, ‘Advances in the study of drawing and handwriting.’, *Human movement science*, vol. 22, no. 2, pp. 131–5, Apr. 2003.

[305] M. M. Schoemaker, C. E. J. Ketelaars, M. van Zonneveld, R. B. Minderaa, and T. Mulder, ‘Deficits in motor control processes involved in production of graphic movements of children with attention-deficit–hyperactivity disorder’, *Developmental Medicine & Child Neurology*, vol. 47, no. 6, pp. 390–395, Jun. 2005.

[306] L. W. Doob, *Hesitation: Impulsivity and reflection. Contributions in psychology, No. 15.* .

[307] K. Jokinen and J. Allwood, ‘Hesitation in Intercultural Communication: Some Observations on Interpreting Shoulder Shrugging’.

[308] M. Corley and O. W. Stewart, ‘Hesitation Disfluencies in Spontaneous Speech: The Meaning of um’, *Language and Linguistics Compass*, vol. 2, no. 4, pp. 589–602, Jul. 2008.

[309] H. Leclercq, Pascale and Edmonds, Amanda and Hilton, *Measuring L2 Proficiency: Perspectives from SLA*. Multilingual Matters, 2014.

[310] D. P. Harrison, Diana and Burkes, Ted M and Seiger, ‘Handwriting examination: Meeting the challenges of science and the law’, *Forensic Science Communications*, vol. 11, no. 4, 2009.

[311] M. C. Fairhurst, L. Smith, J. Mitchell, S. L. Smith, and J. Mitchell, ‘Automated image analysis in visuo-motor testing for the specification of an integrated evaluative and therapy support tool for rehabilitation’, *IEEE Transactions on Rehabilitation Engineering*, vol. 3, no. 1, pp. 103–111, Mar. 1995.

[312] M. C. Fairhurst and S. L. Smith, ‘Application of image analysis to the evaluation of visuo-spatial function in rehabilitation programmes’. p. 2/1-2/3, 1992.

[313] M. M. P. Hoy, M. Y. Egan, and K. P. Feder, ‘A Systematic Review of Interventions to Improve Handwriting’, *Canadian Journal of Occupational Therapy*, vol. 78, no. 1, pp. 13–25, Feb. 2011.

[314] S. N. VENUMADHAVA, GS and ALUR, ‘Investigation of forged documents’, *Reviews of Progress*, vol. 1, no. 5, 2013.

[315] J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, D. Ramos-Castro, and J. Ortega-Garcia, ‘Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems’, *Forensic Science International*, vol. 155, no. 2–3, pp. 126–140, Dec. 2005.

[316] J. Han, M. Kamber, and J. (Computer scientist) Pei, *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann, 2012.

Publications arising from this work

Conference Papers

- T. Islam and M. Fairhurst, "Natural Revocability in Handwritten Signatures to Enhance Biometric Security," *2012 International Conference on Frontiers in Handwriting Recognition*, pp. 791-796, 18-20 September, 2012, Bari, Italy.
- T. Islam and M. Fairhurst, "Investigating an objective measure of writer hesitation for forensic analysis of the handwritten signature," 2016 4th International Conference on Biometrics and Forensics (IWBF), pp. 1-6, 3-4 March, 2016, Limassol, Cyprus.

Appendix A:
Supporting Documentation for the Ethics Approval Procedure

A.1 Participant Information Sheet

PARTICIPANT INFORMATION SHEET

Enhanced signature database collection project

You are being invited to take part in a research study to help us to understand better way people develop and produce their signature style and how this relates to other forms of handwriting. The aim of this project is to build up a database of handwritten data, including the signature, which will be used in research, development and evaluation of automatic handwriting analysis systems and related technologies. Before you decide to participate it is important for you to understand why the data collection is being carried out and what it will involve. Please take the time to read the following information carefully and discuss it with others if you wish. Feel free to ask the researcher if there is anything which is not clear or if you would like more information. Take time to decide whether or not you wish to participate. Thank you for reading this.

Purpose of the study

The overall aim of this project is to establish a database of handwritten samples based on the handwritten signature, but enhanced by the addition of short, simple non-signature handwriting samples. The database will be used by researchers at the University of Kent to develop and improve an understanding of the use of such data in biometric identification tasks and in forensic analysis of handwritten documents.

There are three parts to this study:

Part A: Samples of your signature will be captured using a standard pen of familiar style and feel, and an electronic graphics tablet connected to a computer. The system allows you to write normally on a sheet of paper overlaid on the tablet surface, with the pen movement tracked and a representation of your writing stored in the computer.

Part B: We will ask you to develop a new signature (or we may suggest a signature form for you to use) and we will collect samples (using the same acquisition system as in Part A) of this new signature. This will help us to improve our understanding of how a signing style develops.

Part C: We will ask you provide samples (using the same acquisition system as in Part A) of handwritten examples of the numerals “0” to “9”, and the alphabetic strings representing the months of the year, “January”, February”, etc. Other similar character strings may also be substituted in some sessions. This is to help us to gain greater insight into your personal writing processes, and to investigate whether additional samples can increase our confidence in recognizing a signature as belonging to you.

You may be asked to return and repeat some or all of the data collection process, both to increase the number of available samples per user, and also to help us to understand changes in handwriting styles with time. Not all volunteers will take part in all parts of the collection process.

What will happen to the samples I provide?

The data that you donate will form part of a database which will be owned and maintained by the University of Kent. The data will be used by the University for Research Purposes only.

When you participate, your samples will be stored so that they are linked to a reference number rather than your name. Only the research team collecting the data will be able to link your samples with you personally, and this information will be kept strictly confidential within the research team. It will not be passed to any third party.

Withdrawal

Participation in any part of the collection process is voluntary and you are permitted to withdraw at any time, without giving any reason. You may also withdraw retrospectively and ask that all data relating to you is destroyed.

What will happen to the results of the evaluations using the database?

The results of the evaluation will be documented and are likely to be published in the scientific literature to help others benefit in the future from the knowledge we have gained. However, no participant will be identified individually and no samples will appear in any publication or report which is published without express permission. Copies of any publication will be available via the contact point noted below.

Contacts for further information

Professor Michael Fairhurst
Department of Electronics
University of Kent
Canterbury
Kent CT2 7NT
Email: M.C.Fairhurst@kent.ac.uk

Tasmina Islam
Department of Electronics
University of Kent
Canterbury
Kent CT2 7NT
Email: ti36@kent.ac.uk

You may retain this Information Sheet

THANK YOU FOR TAKING PART IN THIS STUDY

A.2 Consent Form

Acquisition Number:
Subject Identification Number:

CONSENT FORM

Title of Project: Enhanced signature database collection project

Name of Researcher:

Please Tick:

☐ I confirm that I have read and understood the participant information sheet for the above study and have had the opportunity to ask questions.

☐ I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

☐ I understand that no disclosure of my personal details or images will be made without my consent to any party other than researchers within the enhanced signature database collection project team.

☐ I agree to take part in the above study having my signature image captured and provided to be used within the framework of the enhanced signature database collection project.

Name of Participant: ____

Date: ____________________

Signature of Participant: ________________________

Signature of Researcher: ________________________

A.3 Participant Details Sheet

Participant Details

- Please fill the form in BLOCK CAPITALS

SECTION 1: CONTACT DETAILS

1. Title	Mr. ☐ Mrs ☐ Miss ☐ Ms ☐ Other (please state)
2. Contact name	
3. House number and street	
Town	
County	
Postcode	
4. Contact telephone number	

Section 2: Other Details

5. First names	
6. Middle names	
7. Surname	
8. Gender	☐ Male ☐ Female
9. Date of birth	day month year
Or Indicate your age range:	☐ Under 25 ☐ 26-40 ☐ 41-65 ☐ Over 65
10. Nationality	
11. First speaking and writing language	
12. Will you be participating in all parts?	☐ Yes ☐ No
13. Occupation	
14. Did you read the information sheet?	
15. Did you read and sign the consent form?	

Participant Identification number (to be filled in by supervisor):

A.4 Ethics Review Checklist

STMS Faculty

Checklist for Research Projects Involving Human Participation

Project Title: **Enhanced signature database collection project**

Researcher 1: **Professor Michael Fairhurst**

Researcher 2: **Tasmina Islam**

Status: Undergraduate/**Postgraduate**/**Staff**

Does the project involve: Clinical populations? Children (those under 16 years)? Vulnerable adults (e.g. those with mental health problems, learning disabilities, prisoners, young offenders, etc.)?	 Yes/**No** Yes/**No** Yes/**No**
Does the project involve the collection of material that could be considered of a sensitive, personal, biographical, medical or psychological nature?	**Yes**/No
Does the project involve procedures that may reasonably be expected to upset or offend participants (e.g. presentation of unpleasant stimuli, arousal of emotion, etc.)?	Yes/**No**
Will the study be conducted on the internet?	Yes/**No**

☐ I have answered NO to all the above categories and do not consider that this project needs to be submitted for Faculty ethical approval. I have provided a short description of the project below, indicating the involvement of human participant.

√ I have answered YES to at least one of the categories and am submitting an application for Faculty ethics approval. An outline of the project is attached.

Please complete the Faculty ethics forms and refer to the guidance available on-line.

Signature (Researcher 1) **Date**

________________________________ ____________

Signature (Researcher 2) **Date**

________________________________ ____________

A.5 Cover Letter

<u>RE: ENHANCED SIGNATURE DATABASE COLLECTION PROJECT: APPLICATION FOR ETHICAL APPROVAL</u>

Please find attached for your attention:

1. 3 copies of the checklist containing the correct title for this project, which is **'Enhanced signature database collection project'.**
2. 3 copies of the project proposal with the highlighted relevant parts of the project pertaining to data acquisition.

We have attached the project proposal so that we could put into context the scope in which the Data Acquisition work package exercise is being carried out.

We hope this meets your requirements.

Thank you

Professor Michael Fairhurst   Tasmina Islam

---------------------   -------------------------

A.6 Ethics Application

Application Form for Ethical Approval from Research Ethics Group

FOR FACULTY USE ONLY	
Received: ____________________	Date Submitted to Reviewers: __________________
Reviewers: ____________________ ____________________	Review Completed: _______________ _______________
	Researcher(s) Notified: ____________

Submit **three** copies of this form **TO THE FACULTY OFFICE** and attach the following to each form:

- your research proposal
- the participant information sheet
- the participant consent form
- any questionnaires, scales, measures, letters and phone/verbal scripts to be used
- debriefing materials

Name of Investigator: ***Professor Michael Fairhurst*** Email: ***M.C.Fairhurst@kent.ac.uk***
Status: Undergraduate/Postgraduate/***Staff***

Name of Investigator: ***Tasmina Islam*** Email: ***ti36@kent.ac.uk***
Status: Undergraduate/***Postgraduate***/Staff

Project Title: ***Enhanced signature database collection project***

- Indicate here if the proposal is a procedural modification of a previously reviewed project: Yes/***No***

If yes, what was the title of previously reviewed project:

N/A

_

Name of Student/Supervisor in previous project:

N/A

_

List Changes in Current Project: ***N/A***

Source of participants: ***University staff, students and the general public if possible***

Describe the project in no more than one page (summarise the background and hypotheses and detail the procedure to include the conditions experienced by the participants, stimulus, materials and response measures): ***Attached.***

Consent (Please see Consent Checklist)

Is prior informed consent to be obtained? ***Yes***/No

From participants ? ***Yes***/No

Describe the means of obtaining prior consent.

By reading the information sheet and signing the associated consent form with the opportunity to discuss this with a member of the research team.

If prior informed consent is not to be obtained, give reasons:

N/A

Will participants be explicitly informed of what the researcher's role/status is? ***Yes***/No

Will participants be told of the use to which data will be put (e.g. research publication, teaching purposes, media publicity)? ***Yes***/No

Deception
Is there any deception involved? Yes/***No***

If yes, describe the deception and the reasons for its use

N/A

Debriefing
How will participants be debriefed? Written/***Oral***

If they will not be debriefed, give reasons:

N/A

Withdrawal from the investigation
Will participants be told explicitly that they are free to leave the study at any time without jeopardy? ***Yes/***No

When and how will this be done?
In the Participant Information Sheet and associated Consent Form

Confidentiality
Under the Data Protection Act information about a participant is confidential unless otherwise agreed in advance. Will confidentiality be guaranteed? ***Yes***/No
If yes, what steps will be taken to ensure this?

Acquired images will be linked to an identity number and only this identity number will be used in data analysis.
If no, what procedures will be taken in advance of obtaining consent (how will participants be warned)?

N/A

Protection of participants
Are the participants at risk of physical or psychological harm greater than encountered in ordinary life? Yes/***No***

If yes, describe the nature of the risk and steps taken to minimise it:

N/A

Is the information gathered from the participants of a sensitive or personal nature? ***Yes***/No
If yes, describe the procedures to be used for:

a) assuring confidentiality

Identity numbers will be used in place of participants' names in the data analysis and storage. The participants are assured in the consent form that their signature images will only be used within the framework of the 'Enhanced signature database collection project'

b) protecting participants from stress

The image acquisition process will be clearly explained and participants will be free to have breaks, or to terminate the session at any time.

Observational research

If observational research is to be conducted without prior consent, please describe the situation in which observations will take place and say how local cultural values and privacy of individuals will be taken into account.

N/A

I have read the Faculty policies regarding the use of human participants and agree to abide by them. I am also familiar with the ethical principles listed in the Research Ethics Handbook with regard to human participants. I further agree to submit any significant changes in procedures or measurement instruments for additional review.

- Signed:

Researcher(s)

Name: ***Professor Michael Fairhurst*** Signature: ____________________ Date: _________

Name: ***Tasmina Islam*** Signature: ____________________ Date: _________

c) Please remember to attach
- **your research proposal**
- **the participant information sheet**
- **the participant consent form**
- **any questionnaires, scales, measures, letters and phone/verbal scripts to be used**
- **debriefing materials**

Action Taken

☐ Approved

☐ Approved with modifications or conditions noted below

☐ Action deferred. Please supply additional information or clarification noted below.

Date ___________________

Stamped