An assessment of the usability of biometric signature systems using the Human-Biometric Sensor Interaction model

Richard Guest[1], Michael Brockly[2], Stephen Elliott[2], James Scott[1]

[1]School of Engineering and Digital Arts
University of Kent
Jennison Building
CT2 7NT, Canterbury, UK
r.m.guest@kent.ac.uk; jas63@kent.ac.uk

[2]Biometrics Standards, Performance and Assurance Laboratory
Purdue University
155 South Grant Street
47907 West Lafayette, IN, USA
mbrockly@purdue.edu; elliott@purdue.edu

Abstract

Signature biometrics is a widely used form of user authentication. As a behavioural biometric, samples have inherent inconsistencies which must be accounted for within an automated system. Performance deterioration of a tuned biometric software system may be caused by an interaction error with a biometric capture device, however, using conventional error metrics, system and user interaction errors are combined, thereby masking the contribution by each element. In this paper we explore the application of the Human-Biometric Sensor Interaction (HBSI) model to signature as an exemplar of a behavioural biometric. Using observational data collected from a range of subjects, our study shows that usability issues can be identified specific to individual capture device technologies. While most interactions are successful, a range of common interaction errors need to be mitigated by design to reduce overall error rates.

1. Introduction

This study is aimed at understanding the user interaction with biometric implementations, referring to technologies that capture and assess physical or behavioural characteristics of persons, most prominently for purposes of security and identity. Accurately understanding how a user interacts with biometrics in order to maximize the accuracy performance of the system; the user's experience is of critical importance given the recent widespread deployment of systems across large and varied populations (such as in passports and identity cards), and the anticipated growth in usage over the coming decades [1]. Conventional methods of performance evaluation for biometrics focus on error rates in recognition [2] using metrics such as ‘false rejection rate’ and ‘false acceptance rate’ relating to genuine rejection and impostor access respectively. While other statistical measures exist (for example ‘failure to acquire’ which assesses the rate at which the biometric sensor is unable to capture a sample) these metrics provide statistics on how well the overall algorithmic/sensor implementation performs. Although adequate for calculating performance, these metrics do not indicate how this performance was affected by the user interacting with the system [3]. When deployed within a public setting it has been noted that performance of a system drops, not because of a change in the algorithmic implementation, but purely due to the nature of the user interaction with the technology not adhering to the expected methodology, typically due to unfamiliarity with the system, poor user instructions, positioning with respect to the sensor resulting in poor sample capture, or a combination of these factors [4]. With frontline deployment of what might be described as immature technology, accurate assessment of user interaction is timely and of paramount importance as a research issue [5].

In an attempt to assess, understand and react to biometric systems usability issues, the HBSI model was developed [6] to present a user-centric assessment of performance, enabling usage errors to be decomposed and attributed to a multitude of factors including incorrect user interaction, performance modification due to ergonomics and user interface, and error in sampling/poor quality of sample, alongside conventional measurements of algorithmic performance. Adopting this model it is possible for developers, integrators and end-users to pinpoint exactly where usage error occurs. The original HBSI model was developed for biometric fingerprint collection, a modality which relies on the instantaneous capture of a fingerprint image. This process, also used in other physiological modalities such as iris, hand geometry and face, relies upon performance and usability assessments at a single moment in time within the HBSI model. While further studies have demonstrated the agility of the model within an instantaneous context [7], the original model did not allow for the analysis of temporal/behavioural biometric modalities – systems which capture and analyse a sequence of events to prove identity, examples of which include movement analysis and dynamic signature. Furthermore many deployed biometric implementations use multiple modalities alongside other non-biometric user interaction requirements (in so called complex systems), such as the presentation of a token or entering of a personal identification number (PIN), all of which require an overall temporal assessment of a system’s usage. An example of this is the recently introduced automated biometric passport gates found at numerous border control posts. Here the user approaches a kiosk, enters the passport into a reader, has a facial image captured, retrieves the passport and passes through (if successful). Although the primary biometric (face) is physiological, the overall interaction is behavioural.

Dynamic signature systems are a widely adopted behavioural biometric that have shown to perform well in comparison against other biometric systems [8]. The signature modality is widely accepted by society as a legally admissible form of authentication and because of this, the modality has particularly high take-up within the financial, legal and commercial transaction sectors. As a behavioural biometric, identity is attributed on how a subject signs a signature based both on the final signature 'image' but also the temporal constructional aspects of production. Alongside other behavioural biometric systems a signature sample will have inherent variability, which partially explains higher error rates than other physiological biometrics such as iris or face. In assessing usability and signer interaction with a signature capture device it is important to consider interactions with the device from the start of signing until the end of the entire capture process and incorporate any modifications to the signing process and interaction errors with the capture device.

In this paper we will explore the application of the HBSI model to the dynamic signature modality. By investigating how subjects use a range of common signature technologies to enrol and verify, we explore how and why interaction errors occur and use the HBSI to metricize both system and user related performance errors. The analyses enable a thorough understanding of where common errors are introduced within signature systems and enhance the granularity of attributing errors. In turn, this analysis can assist in the design of signature systems from software, device ergonomics and user instruction point of view.

2. Signature Systems and Interaction Assessment

As discussed, biometric systems for human identification can be broadly defined into two categories: physiological biometrics related to physical characteristic of a person (for example, facial image, fingerprint, iris pattern) and behavioural biometrics related to how a person performs an action (for example walking pattern/gait). In general physiological biometrics can be captured in an instant whereas behavioural biometrics require a capture period of a number of samples resulting in a single presentation over a (short) period of time. In assessing usability aspects of biometric systems it's therefore critical to establish interaction levels appropriate to the nature of the technology. With respect to behavioural biometrics, this means assessing interaction across the entire capture period.

Biometric identification through signature production is an example of a behavioural biometric in that recognition processes analyse how a subject produces a signature. Signature assessment is performed either on a 'static' basis, by examining the completed signature image (the conventional point-of-sale scenario when comparing a signature on the reverse of a credit card to that donated at the payment counter), or on a ‘dynamic’ basis where temporal aspects of the signature performance are assessed including pen positions, velocities, and other metrics.

Static signatures can be captured using a conventional scanner, while dynamic signature require the use of a specialized digitising/tablet device that is able to capture pen status information as a signer writes on a surface using a stylus. With these devices, stylus position, pressure and tilt are captured alongside a time offset during signature capture. These devices may or may not provide ‘ink’ feedback to the signer (either as virtual ink on a back projected tablet surface or actual physical ink from a pen writing on an overlaid sheet of paper) during this process. Tablet devices using finger motion interaction have become increasingly ubiquitous with devices running operating systems such as iOS and Android, revolutionizing interaction design. Future developments for signature assessment may encompass finger-based singing alongside the conventional stylus input method.

A number of studies have examined different signature technologies and their effects on signing capabilities. Many aspects of the sensor design will affect the user acceptance as well as overall usability. Some factors that can affect the way a user interactions with the sensor include the size, shape, grip, and stylus of the device [9]. Users also inherently prefer the biometric device to be at a certain height level [10]. If the device is not at an optimal height or angle for user comfort, additional errors may occur as the user simply cannot easily use the system. As the market changes and there is a shift towards mobile systems on tablets and smartphones, additional factors such as lighting, screen sensitivity and weight will impact the usability of the device. These factors, along with their usage on different operating systems will create a need for greater interoperability of signatures.

Interoperability in signature systems is a two-fold process. The first type of interoperability occurs with the end signature result. Different devices need to be able to yield the same signature so that multiple devices can be used without any loss of performance. Interoperability also refers to the interaction procedure between the human and the device. Different systems should follow the same steps and procedures in order to prevent additional user systems that are unique to only specific devices. HBSI errors may sometimes be unique to specific devices [11] and will vary greatly between mobile and non-mobile environments.

2.1 The Human-Biometric Sensor Interaction

The evaluation method for this study originates from traditional metrics used to measure biometric system performance. These metrics include failure to enrol (FTE) and failure to acquire (FTA). The metrics are assigned as the end result of a biometric sample collection process, simply determining that some data collection error occurred that could be due to either the system or the user. HBSI metrics ask the question “What occurred during the data collection procedure to cause an FTE or FTA?” In order to answer this question, the entire sample collection procedure needs to be analysed, instead of just the end result. Video analysis for error coding is commonly used to determine what the error was caused by and when it occurred [12], [13]. In order to determine the error type, the HBSI methodology is applied.

The concept for the Human-Biometric Sensor Interaction (HBSI) model was originally developed in 2004 by researchers at Purdue University [14]. The model consists of multiple components that come together when a human uses a biometric system. Overlaps between each component are also essential when conducting an evaluation of a biometric system. The evaluation metrics of HBSI include ergonomics, usability and sample quality. Figure 1 shows how the elements of a biometric implementation and the associated evaluation characteristics interact, with the extractable metrics from HBSI forming the intersection between all components.

Research into HBSI has progressed beyond original evaluations into an error framework shown in Figure 2. The HBSI Error Framework is used to classify a presentation made on behalf of the user to a biometric system as an error or a successfully processed sample (SPS). These classifications are determined based on presentations made to the system that are either correct or incorrect.

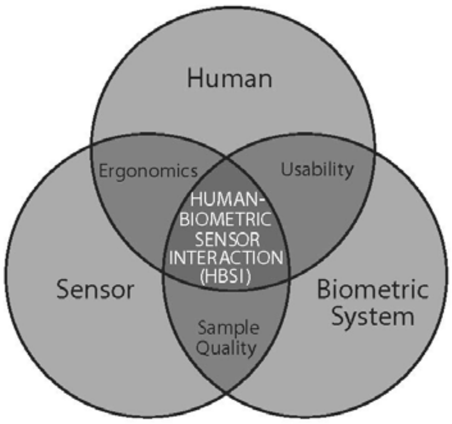

Figure 1: HBSI Model [10]

Determining the root cause of the error is essential, as researchers and businesses alike are interested in determining if it is due to a system error, a user presentation error, or a combination of the two. The five error metrics and one success metric are defined as follows:

- Defective Interaction (DI) – An incorrect presentation that is not detected by the biometric system.

- Concealed Interaction (CI) – An incorrect presentation that is detected by the biometric system but is classified incorrectly. The system accepts the presentation and the sample proceeds onto the general biometric model.
- False Interaction (FI) – An incorrect presentation that is detected by the biometric system and classified correctly. This incorrect presentation is correctly handled by the system, rejecting the sample and providing feedback to the user.
- Failure to Detect (FTD) – An FTD is a correct presentation that is not detected by the biometric system.
- Failure to Process (FTP) – This correct presentation is detected by the system but due to system errors, is not processed or saved successfully.
- Successfully Processed Sample (SPS) – A correct presentation that is both detected and processed correctly by the biometric system.

These metrics have been successfully used for the assessment of fingerprint systems and have shown that clear distinctions of systems-related and user-related errors can be made within the analysis of system performance [6], [15].

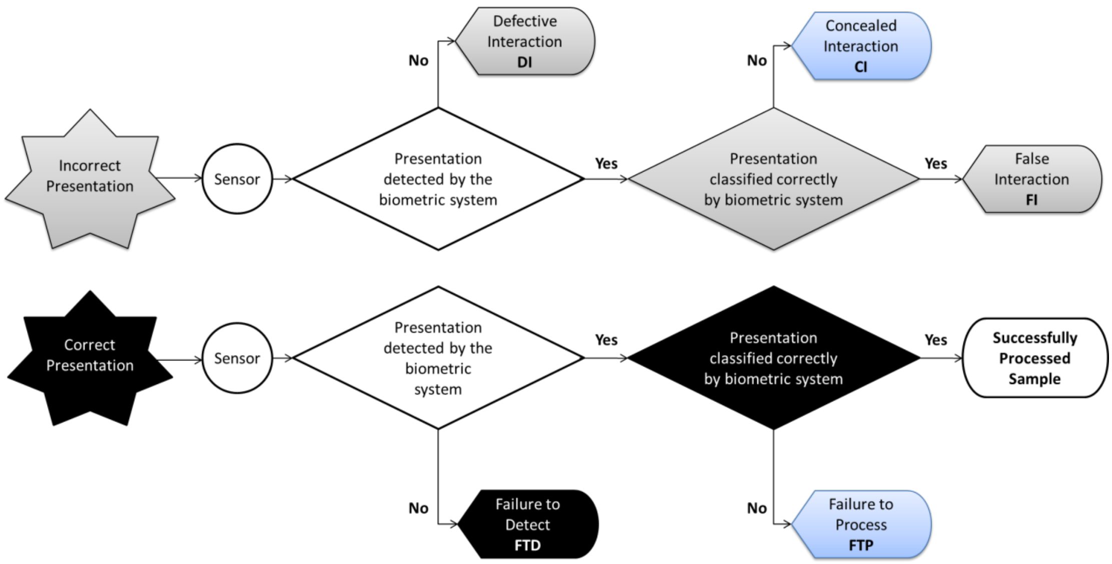

Figure 2: HBSI Error Framework [11]

3. The application of HBSI to Dynamic Signature Systems

Applying dynamic signature to the HBSI model, the individual error metrics of this modality accurately map to the error framework. The potential errors that can occur in Dynamic Signature Verification (DSV) are as follows:

- DI – The signature is deemed incorrect and no signature channel data is collected. Due to the incorrect presentation, the signature is not detected by the system. An example of this can occur if an incorrect stylus is used, causing the system to not detect the signature.
- CI – The detected signature, although incorrect, is not identified by the biometric system as an error, and can subsequently be used within an enrolment or verification operation. This is the most dangerous type of error because it goes unnoticed by the system and accepted as an SPS. A successful signature forgery will result in a CI.
- FI – The incorrect signature presentation is properly deemed erroneous and not allowed into the system. The signature may be deemed unrepresentative by the signer so they will hit the 'clear' button, or the system will correctly reject a forgery.
- FTD – The correctly-presented signature is not detected by the biometric system. This may occur due to temporal data recording issues, or possible system failure.
- FTP – The signer has donated a correct signature that was detected by the system but a system error causes the signature to be unrepresentative of the subject, or simply fail to store it to the database. This may be caused by insufficient system memory or latency issues.

As a temporal/behavioural biometric there are a number of additional considerations when applying HBSI to an interaction analysis. As samples are collected and analysed over a time period, rather an instance, it is possible for a subject to modify their donation behaviour during the capture period. For example, a signer might start signing their signature, reject the (partially completed) signature during the capture and then start the signature again. If the restarted signature is accepted by the signer then, in the context of the HBSI model, a single capture session has resulted in both an incorrect and correct presentation.

This user-referred modification of behaviour is not unique to the signature modality as it can be observed in other behavioural modalities (for example gait). In applying the HBSI model we should strike a balance between ensuring meaningful application of the analysis and, in as far as is possible, applying a generic framework across all biometrics (physiological and behavioural). To enable a sub-division of the capture process into abortive/incorrect and correct attempts we must first consider a harmonized set of definitions that enables the granulation of the capture process.

3.1 Terminology

Conflicting terminology in the field of biometrics exists in many different sources [16–19]. It is paramount to establish a set of universal terms in order to understand biometric processes across modalities. This research uses the following terms in describing the biometric capture process:

interaction, presentation, attempt, and transaction. Previously validated for hand geometry in [20], this terminology can also be applied to signature. Figure 3 shows how these elements interact within a signature system.

A transaction is defined as "the sequence of attempts to the system on the part of the user for the purpose of enrolment, verification or identification" [17]. An enrolment transaction includes all of the signatures donated by the user in order to create a biometric template. A verification transaction consists of all of the signatures donated by the user in effort to verify against their template.

An attempt is defined as "the submission of one (or a sequence of) biometric samples to the system on the part of the user" [17]. In a DSV enrolment, it is common that one enrolment attempt will consist of multiple signatures (usually between 3 and 5). If the user is unable to enrol into the system in 3 signatures, they will need to begin another attempt to enrol successfully. A transaction may consist of multiple attempts (i.e. enrolment or verification over time-separated transactions). Most signature enrolment and verification transactions will consist of a single attempt.

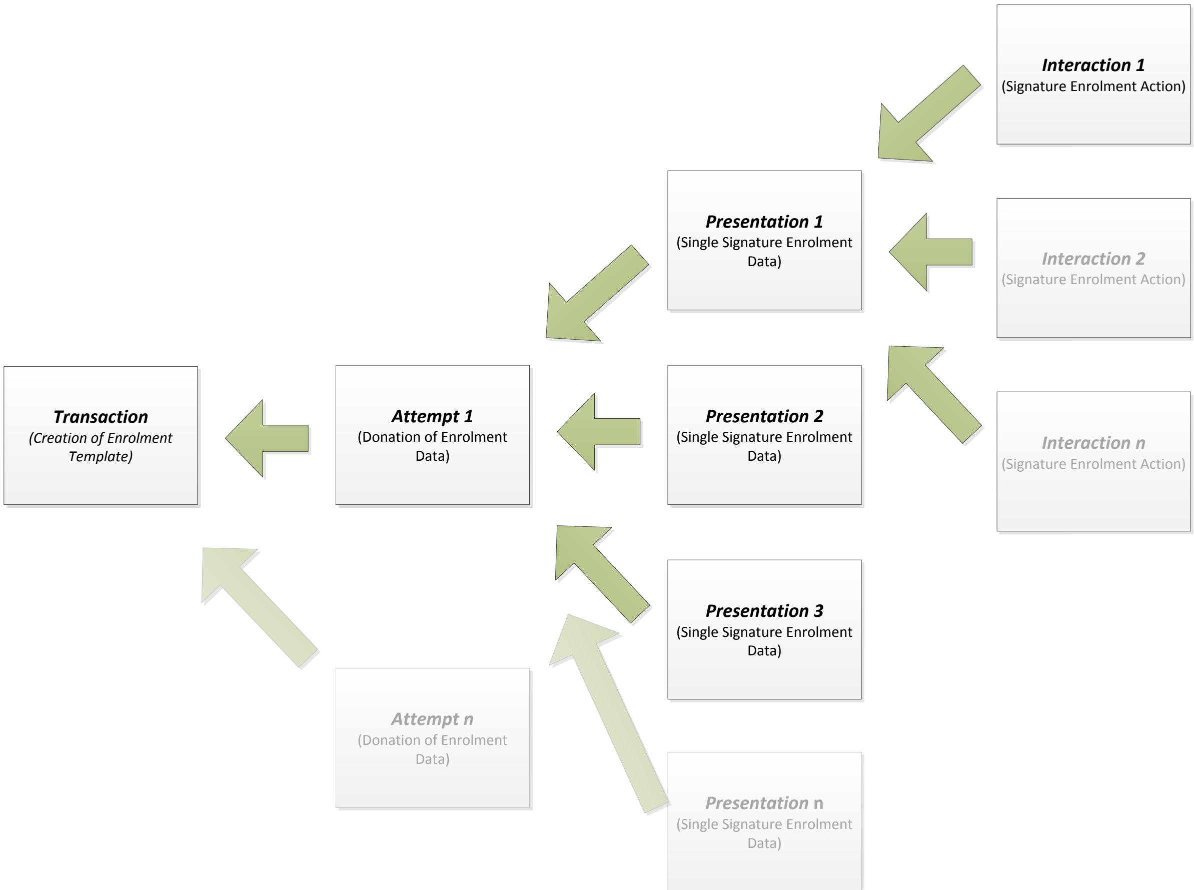

Figure 3: Capture terminology applied to dynamic signatures

The three signatures constituting an attempt are considered to be three separate presentations. A presentation is defined as "the submission of a single biometric sample to the system on the part of the user" [17]. Each time the user signs their signature, for means of enrolment or verification, will create a presentation of their signature to the biometric system on their behalf.

Each presentation is made up of any number of interactions. An interaction is defined as "the action(s) that take place within a presentation" [20]. Interactions refer to all individual actions performed by the user during the signing process; such as creating the signature by interacting with the pen to write on the digitizer's surface, or other interaction affecting the signature outcome such as accidentally placing their hand on the pressure pad. If successfully executed, a presentation may consist of a single interaction. Conversely, may interactions may exist within a presentation, which can either end with a successful or unsuccessful presentation within the HBSI model.

The important issue here is addressing multiple interactions within a single presentation. The divide between an interaction and a presentation is necessary in order to understand how interaction errors affected the overall signature presentation. Each individual signature presentation is made up of many small interactions which may be correctly or incorrectly executed. Depending on the impact of an incorrect interaction, it may or may not result in the total presentation as being classified as correct or incorrect.

3.2 User Reset (UR) State

Importantly in the context of behavioural biometrics it is possible for a subject to restart or abandon a presentation. Within our four-stage terminology model each restart would be deemed a new interaction. To capture the behaviour of this possibility, we define a new state within the HBSI model – User Reset (UR). This is independent of whether a presentation is correct or incorrect and can occur at any point during donation.

Within the dynamic signature modality, consider the scenario where a signer starts to sign correctly but is not satisfied with the signature being produced. The signer may be given an option to reset the signature (by clearing the screen and/or restarting the capture process) within a presentation procedure. The first interaction would be halted with an UR outcome, and a second interaction would then proceed which may result in a correct or incorrect presentation (or another UR). The application of the HBSI model to signatures therefore has outcomes of either a correct or incorrect presentation. An interaction may end with an UR whereby a new interaction starts and the presentation recording is reset. Apart from this minor addition to HBSI for behavioural modalities (including signature), the generic HBSI model previously devised is directly applicable.

4. An assessment of signature systems using HBSI

To assess the application of the revised HBSI model to signature modality systems, a series of interaction videos were recorded of subjects using a common signature enrolment and verification scenario. Twenty-six subjects were asked to enrol and then verify on a commercial signature verification engine. Subjects used two common tablet devices to capture signatures. Table 1 details these two devices. The first device employed a back-projected screen that displayed virtual ink as the test subject wrote on the tablet surface. Driven by the commercial verification software, this screen also contained an area for user interaction allowing a signer to confirm, reset or cancel as signature. The second device used a capture area that did not provide inking feedback as the user signed with an inkless pen (denoted non-inking). The signer could, however, see their signature being drawn during the signing process on a screen located on the same table as the inkless capture device. This study was denoted non-inking study 1. To analyse the interaction with the devices all sessions were recorded using two cameras mounted from the side and above the signature tablet.

A series of performance statistics were extracted from the experiments. Alongside conventional enrolment and verification performance rates (including FTE and FTA) that show the performance of the engine, we also manually analysed the videos by encoding a series of interaction errors in the subjects' signature donation sessions. Applying the HBSI model allowed individual responses to be classified as correct or incorrect presentations. Further to this, individual correctly presented samples were categorized as leading to an SPS or FTP. Likewise, incorrectly presented samples were also categorized leading to an FI or CI.

To further investigate the use of the non-backlit digitizer (the most commonly deployed sensor) a further study was conducted using the same software (denoted non-inking study 2). This collection included 56 subjects, enrolling and then subsequently verifying against their template. Enrolments consisted of 3 signatures, and then the subjects had one attempt to verify on the system. For this second study, an operator controlled the donation process and was able to react to a request to restart/clear the signature, whereas in the first study the user could see their signature construction on a desk-mounted screen.

Table 1: Sensor Specifications

Sensor Type	Sample Rate (Hz)	Pressure Levels	Movement Resolution (dpi)	Movement Detection Technology	Platen Size (mm)
Non-inking	100	0-127	300	Semiconductive	88.9 x 53.1
Virtual ink	100	0-127	300	Resistive	76.0 x 56.0

In analysing the results the aims across all the studies documented in this paper were to a) investigate the performance differences between the major capture technologies and b) assess where these performance differences were caused by human or system based interaction issues.

5. Results

As an initial assessment it is possible to examine both devices across the three trials using the conventional system performance metrics exploring both the enrolment and verification statistics separately. Table 2 shows the enrolment statistics detailing the number of subjects and interactions within each trial. The larger number of interactions to subjects is indicative of the multiple attempts made to successfully donate three signatures. As can be seen, the majority of interactions were successful with most subjects being able to enrol on the first attempt. Only two subjects within each of the virtual ink and non-inking study 1 trials were unable to enrol on the system.

These results show that although most subjects eventually able to enrol on the system using either capture technology, there are a large number of unsuccessful attempts. Conventional assessment would group together systems and user errors without exploring the reasons for these errors.

Exploring the difference between non-inking studies 1 and 2 it is interesting to note the broad similarities between results. In study 2, a larger number of subjects were able to enrol at the first attempt with fewer unsuccessful enrolments (and no FTEs). This indicates that even though the hardware and software are a constant, the variables in the process are the cohort and the operator instructions provided to the cohort. If we consider both of these to be similar in composition it indicates the natural range of variability within a common scenario. This highlights the requirement to ensure that errors are correctly attributed to either system or user at each implementation, rather than assigning a global performance metric.

Table 2: Enrolment statistics - conventional assessment

Enrolment Statistics	**Virtual ink**		**Non-inking Study 1**		**Non-inking Study 2**	
Total attempts	45		36		62	
Total subjects	26		28		56	
Total unsuccessful (enrols)	20	44.44%	7	19.44%	2	3.23%
Total successful (enrols)	24	53.33%	27	75.00%	56	90.32%
Abandoned attempts	1	2.22%	2	5.56%	4	6.45%
Enrolled first time (subjects)	18	69.23%	22	78.57%	54	96.43%
Enrolled on 2nd attempt (subjects)	3	11.54%	3	10.71%	2	3.57%
Enrolled on 3rd attempt (subjects)	2	7.69%	0	0.00%	0	0.00%
Enrolled on 4th attempt (subjects)	1	3.85%	0	0.00%	0	0.00%
FTE (subjects)	2	7.69%	2	7.14%	0	0.00%

Extending the conventional assessment to the verification results (Table 3) shows a marked contrast in the difference between the two technologies. In the virtual ink implementation the majority of successfully enrolled subjects were consequently unable to verify either as a first attempt or after multiple attempts. Both non-inking trials reversed this finding with most subjects being able to verify which supports the validity of signatures as a reliable biometric (it should be noted that non-inking study 2 only gave subjects one attempt to verify). However it is clear that by examining the total number of unsuccessful attempts there is an issue to explore into why these errors are occurring and precisely how many can be attributed to user interaction.

Table 3: Verification statistics - conventional assessment

Verification Statistics	**Virtual ink**		**Non-inking Study 1**		**Non-inking Study 2**	
Total attempts	50		61		55	
Total subjects	23		27		55	
Total unsuccessful (verify – multiple attempts)	30	60.00%	23	37.70%	-	-
Total successful (verify – multiple attempts)	20	40.00%	37	60.66%	-	-
Total unsuccessful (verify – first attempt only)	12	52.17%	17	62.96%	6	10.91%
Total successful (verify – first attempt only)	11	47.83%	10	37.04%	49	89.09%

Having assessed the conventional statistics which do not indicated where an error has occurred (only that an error has occurred) it is possible through the use of HBSI to assess why sub-optimal performance occurs. Table 4 shows the HBSI error states across the enrolment interactions, focusing explicitly on how errors are divided between user and systems. The first two lines of the table (labelled 'CP' and 'IP') show that the majority of presentations across all input devices are correct indicating that the remaining error is due to software misclassification of inconsistent samples. Performance between capture devices was fairly uniform. Typically around 10% of the presentations contained a user-introduced error (IP). Examining the distribution of HBSI error states for the virtual-ink it can be seen that within the correct presentations (shaded in grey) a large percentage of errors (40%) were caused by a failure to process (FTP) – subjects here had successfully donated a signature but the system returned a non-match. This indicates that the virtual-ink capture technology introduced larger than normal variations within signature production. Usability errors within the virtual ink enrolment samples showed that the few samples that were incorrect presentations there was a mix between CI and FI. This shows that, in general, users didn't have usability issues with the device even though it caused variation in production performance. Assessing the enrolment data from non-inking tablet studies it is evident that this tablet resulted in a larger number of SPS samples, even though roughly the same percentage of samples were CPs. This indicates a greater stability of signature data even without an ink feedback mechanism. Considerably fewer correct samples were misclassified (FTD/FTP) by the signature engine. Although few in number, all IPs were correctly identified as such by the system. A larger number of URs were also noted, particularly within study 2 where an operator controlled the reset operation. With respect to this in study 2 the operator could reset the signature without comparison while in study 1 a reset could have been called following a visual inspection of the remote image.

Table 4: HBSI enrolment statistics

Enrolment HBSI Statistics	Virtual ink		Non-inking Study 1		Non-inking Study 2	
CP	39	86.67%	31	86.11%	58	93.55%
IP	6	13.33%	5	13.89%	4	6.45%
SPS	21	46.67%	27	75.00%	56	90.32%
FTP	18	40.00%	4	11.11%	1	1.61%
FTD	0	0.00%	0	0.00%	1	1.61%
FI	2	4.44%	3	8.33%	0	0.00%
CI	3	6.67%	0	0.00%	0	0.00%
UR	1	2.22%	2	5.56%	4	6.45%

Table 5 shows these verification statistics when just considering the first attempt at verification. Only subjects that were able to enrol of the system were able to subsequently verify. Again it can be seen that the vast majority of presentations were correct (CP) across the three studies and the rates of the software successfully processing these very considerably, broadly in line with the enrolment statistics. Therefore our conclusions are that the virtual-ink capture technology introduced larger variations within signature production than for non-inking devices. No subjects generated UR requests within the verification process, indicating a habituation of process – 'my signature has already be accepted in enrolment therefore I don't need to reject any further'. This may not be a valid hypothesis for the non-inking system where a larger number of IPs are present.

Table 6 shows the results from the verification presentations across the devices and studies when subsequent attempts are made to verify following a failure to verify (the results also including the first attempt). In each of the studies two (disjoint) subjects failed to enrol after six attempts and were deemed as FTE. Non-inking study 2 results are not included in this table as subjects only were given one chance to verify.

Table 5: First attempt verification statistics

Verification Statistics	Virtual ink		Non-inking Study 1		Non-inking Study 2	
CP	22	95.65%	22	81.48%	53	96.36%
IP	1	4.35%	5	18.52%	2	3.64%
SPS	11	47.83%	17	62.96%	49	89.09%
FTP	11	47.83%	5	18.52%	4	7.27%
FTD	0	0.00%	0	0.00%	0	0.00%
FI	1	4.35%	5	18.52%	2	3.64%
CI	0	0.00%	0	0.00%	0	0.00%
UR	0	0.00%	0	0.00%	0	0.00%

Table 6: HBSI verification statistics

Verification Statistics	Virtual ink		Non-inking Study 1	
CP	47	94.00%	44	72.13%
IP	3	6.00%	17	27.87%
SPS	20	40.00%	23	37.70%
FTP	27	54.00%	21	34.43%
FTD	0	0.00%	0	0.00%
FI	3	6.00%	17	27.87%
CI	0	0.00%	0	0.00%
UR	0	0.00%	0	0.00%

The analysis of the enrolment and verification results enable a more detailed inspection of the where errors were caused by incorrect presentations, and where these exist, where a biometric software system is able to accurately detect these. Our framework also includes the possibility that a correct presentation is not detected at all by the system.

Focussing on incorrect presentations, the video-based analysis across the studies also allows an assessment of what caused the errors that contributed to an IP. The following errors occurred on the back projection with ink feedback:

- Ink dispersed within the OK or Clear button – the software driving the back-projected tablet displayed a background image on the screen which contained a number of 'buttons' which the user could select a signature clear/reset, exit/abandon capture and an 'OK' button to signify successful signature competition. An error occasionally occurring during the an enrolment phase is that the user places the pen within the button area but the system stores this erroneously as an ink point rather than a user interaction (3 instances during enrolment).

- Repeating parts of signature – The signer repeated parts of the signature due to ink feedback errors. This usually occurs when the pressure is not sufficient to register as a pen-down event or a latency error causes a delay in ink deposition (1 instance during verification).
- Non-dispersal of ink – as above the ink is not deposited due to a system or user-pressure interaction error. The user doesn't choose to repeat the missing parts of the signature (2 instances during enrolment, 2 instances during verification).
- Hand on pressure pad – The signer touched the tablet with either a palm or finger causing an erroneous 'pen' record event. The occurrence of this depends on the topology of the capture technology as some tablets will only record events drawn by an induction pen (1 instance during enrolment)

Using our new UR state, this can be (self) triggered by two events:

- Clear button pressed – the signer pressed the clear button and restarted a signature.
- User unhappy with signature – User abandons the signature during the signing process (1 instance during verification).

On the non-inking tablet, the range of interaction errors included instances where the user introduced an erroneous event but, due to the lack of feedback, is unable to respond. The range of errors for the non-inking device includes:

- Hand on pressure pad – The signer touched the tablet with either a palm or finger causing an erroneous 'pen' record. The subject may not know that they did it without seeing feedback (Study 1 – 5 instances during enrolment, 2 instances during verification, Study 2 – 2 instances during enrolment, 1 instance during verification).
- Ink dispersed within the OK or Clear button – although these buttons aren't visible to the signer they exist within the software interface. It is therefore possible for a signer to draw within the button area (Study 1 –2 instances during verification, Study 2 – 2 instances during verification).
- Repeating parts of signature – The signer repeated parts of the signature due to uncertainty when signing – again hampered by non-visual feedback (Study 2 – 1 instance during enrolment).
- Pen pressure not registered/dispersed correctly – the signer used a pen pressure that was not detected by the tablet. With this capture technology the signer has no way or knowing whether an erroneous donation has occurred (Study 1 – 2 instances during enrolment, 16 instances during verification, Study 2 – 2 instances during enrolment).

A UR state can be called if the user has an option to review their signature, for example on a remote screen. This was an option within the non-inking study 1 directly controlled by the signer whilst in study 2 the signer was able to interact with an operator to enable a reset. (Study 1 – 2 instances during enrolment, Study 2 - 4 instances during enrolment).

Generically, there is also the possibility of a system (latency) error which means that nothing is captured during the signing process.

Although there are a number of similarities in the types of user-errors found within the two devices there are a number of key differences that can be considered when designing a signature system.

- Problems in ink dispersal within the virtual ink system causes problems with users abandoning presentations (UR) and modifying through repetition.
- Hands/fingers on the tablet surface causes major issues on technologies that detect pressure (rather than by induction coil pen). This was more prevalent on the non-feedback devices where a signer would be unaware of the production of an erroneous event. Ideally x and y coordinate streams should be filtered to remove sudden movements that are out of scope for normal signature production.
- Incorporating virtual buttons, which may be visually the best solution causes issues a) when a signer accidentally places the pen in the button zone as part of the signature – triggering an 'accept' or 'cancellation' event or b) an attempt to press a button is recorded as part of the signature. Special consideration should be given to the distance between the sizing zone and virtual buttons, or removing the virtual buttons altogether, with 'accept' or 'cancellation' events triggers by other mechanisms.

Assessing these erroneous events it is possible to map a flow through the process of collecting a signature on different capture technologies. Figure 4 and 5 define the capture process flow of the virtual ink and non-inking capture technologies respectively. These charts define the sequence of events within an interaction with each of the decision points denoted with an identifier. By exploring the HBSI events that lead from each of these decisions it is possible to tie each to a particular outcome. These are shown in Table 7 and Table 8 and serve as a reference for the HBSI errors that can occur.

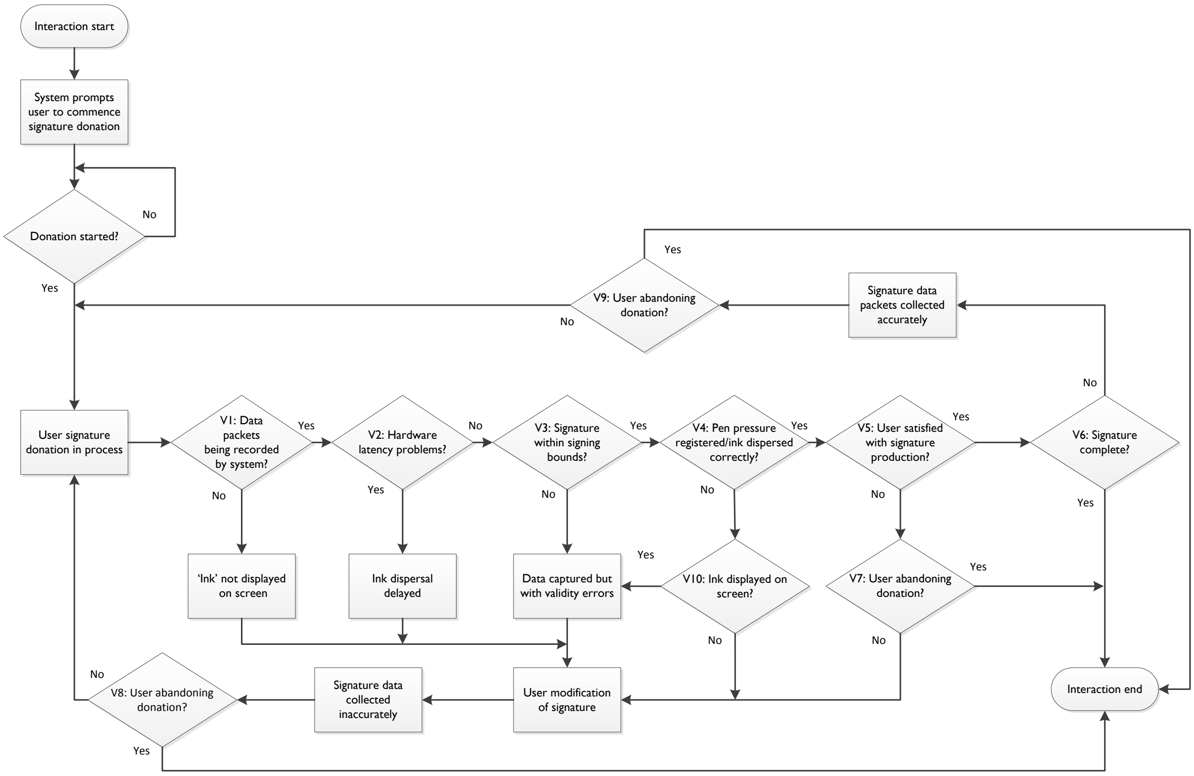

Figure 4: Virtual Ink Process Flow

Table 7: Virtual ink process-to-HBSI mapping

Decision Point	Action	Outcome
V1	No	If CP - FTD. If IP - DI. UR if subsequently abandoned
V2	Yes	FTP if feedback detached from sample donation pen movement (i.e. a CP event)[*]. UR if subsequently abandoned, else CI if successfully enrolled/verified, FI if rejected
V3	No	UR if subsequently abandoned, else CI if successfully enrolled/verified, FI if rejected
V4	No	UR if subsequently abandoned. CI if successfully enrolled/verified, FI if rejected
V5	No	To V7
V6	Yes	SPS
V7	Yes	UR
V7	No	CI if successfully enrolled/verified, FI if rejected. Assumes that an IP has been donated to route through the V5 decision box
V8	Yes	UR
V9	Yes	UR
V10		Inheriting outcome from V4

* Latency issues may be caused by either a delay prior to recording the sample data and displaying the ink (*recording delay*) or after data has been recorded (*display delay*). We assume that if a display delay has occurred, a CP has been donated, whereas a recording delay would result in an IP.

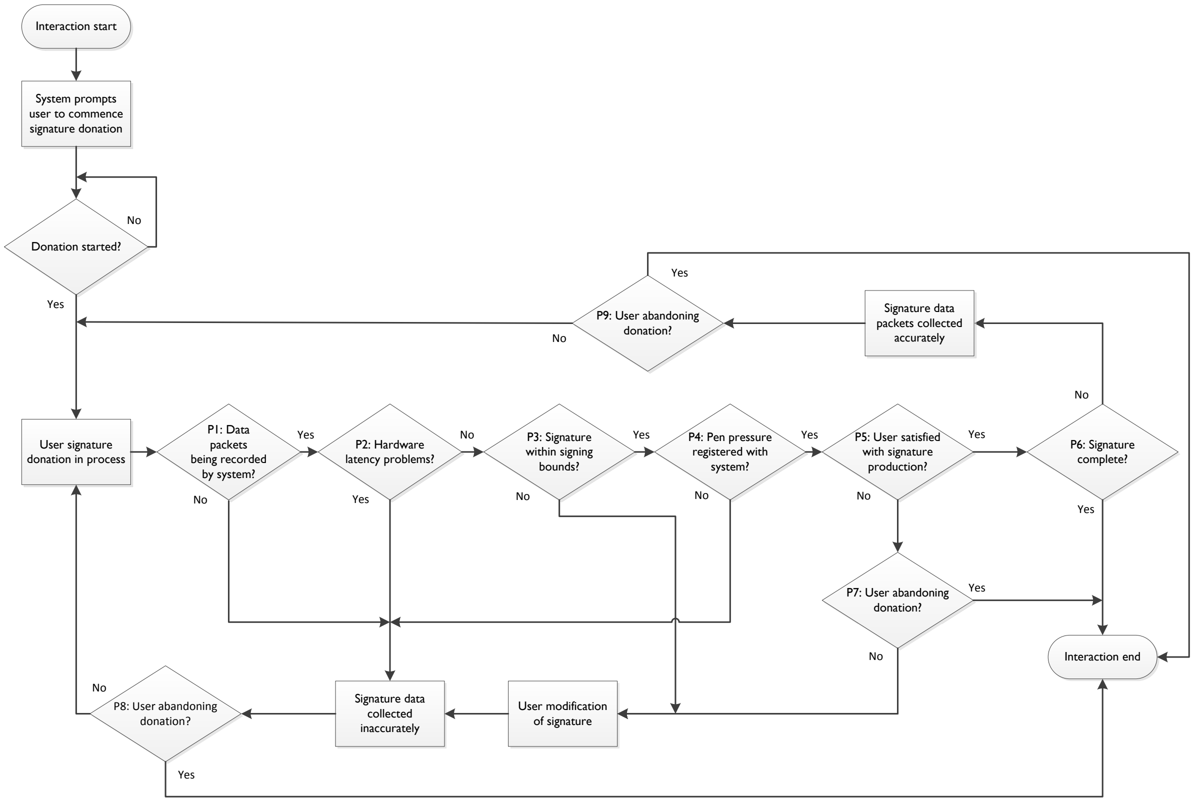

Figure 5: Non-inking System Process Flow

Table 8: Non-inking process-to-HBSI mapping

Decision Point	Action	Outcome
P1	No	If CP - FTD. If IP - DI. UR if abandoned
P2	Yes	If CP - FTP. If IP - CI if successfully enrolled/verified, FI if rejected. UR if abandoned
P3	No	CI if successfully enrolled/verified, FI if rejected. UR if abandoned
P4	No	If CP - FTP. If IP - CI if successfully enrolled/verified, FI if rejected. UR if abandoned
P5	No	To P8
P5	Yes	SPS
P6	Yes	UR
P6	No	CI if successfully enrolled/verified, FI if rejected. Assumes that an IP has been donated to route through the P5 decision box
P7	Yes	UR
P8	Yes	UR
P9	Yes	UR

6. Discussion and Conclusions

In this paper we have analysed the performance of automatic biometric signature systems both in terms of overall system performance and interaction usability. Using the HBSI framework and modifying this to include a new state of UR to include the condition whereby a subject resets their donation process, we have demonstrated the ability of the analysis framework to pin-point where errors occur within the signature modality.

It is encouraging to note the high performance of the signature systems, including the large number of samples that are donated without any usability error. We have, however, noted a number of common issues that can occur and need to be accounted for. It was evident that having 'Reset' and 'OK' buttons within the signing area caused a problem, as did the user touching the signing surface during capture, thereby depositing erroneous ink. The other major error was the non-deposition of ink cause either by latency issues or by insufficient pressure being applied to the tablet surface. These problems can be solved through the design of appropriate on-screen user interfaces and hardware surrounds to protect the signing surface from additional finger interaction.

It is important to remember that, within the process of signing, the element of ceremony plays an important role with respect to the quality of provided signature. If the signature signifies an element of high importance (for example, on a legal document) the user typically signs with greater care, striving for enhanced quality and clarity. This must be considered when applying a signature system to a particular scenario – are there likely to be more URs due to the signer ensuring that the signature is correct or is process speed of essence?

One effect that wasn't noticed in this current trial but that can affect the outcome of temporally stored data is a signer embellishing a signature with an underline or a character modification following the completion of the normal signature. This would be collected at the end of temporal data and would thus cause a potential issue for dynamic systems where two sequences are compared. Embellishing a signature after previous enrolment will also incur an incorrect presentation in the HBSI methodology.

Although this study was designed as a pilot to assess proof of concept in applying HBSI to signature systems (and behavioural biometrics more widely) we have shown clear indicators of how the framework reveals a finer granularity in assigning errors that can be utilised within the design and assessment of biometric systems (signature or otherwise). Within the signature modality, future work will focus on the testing on other types of digitizers – although the two main technologies have been assessed in the current trial, both were assessed within a desk-bound environment. The growth of mobile devices and scenarios leads to a natural extension of this work. We also aim to calculate usability errors automatically from video processes and systems logs, thereby increasing the immediate practicality of the application of this method.

Whilst we have used signature systems as an exemplar of a behavioural biometric, other biometric modalities in the genre (for example, gait [21] and speech) can be assessed using the same method. With the addition of the UR state, subject interactions can be labelled as either incorrect or correct presentations, with the option to abort and reset the capture operation. This process is common to all behavioural biometrics and highlights the generic nature of the framework. The UR state is also found in some physiological biometrics, such as a hand geometry system which uses a PIN to access the user's template.

The HBSI methodology has adapted well to the behavioural modality of signature, without needing any major redesigns. In order to complete the model for universal adaptation, it still needs to be tested in more modalities. Work will continue to explore the application of the HBSI methodology to other behavioural modalities or complex systems involving temporal biometric processes and token presentations. Further research into this field will help to strengthen and improve this model.

References

[1] Acuity Market Intelligence, "The Future of Biometrics Report: 2009 Revised Edition," 2009.

[2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[3] M. E. Schuckers, *Computational Methods in Biometric Authentication*. London: Springer London, 2010, p. 317.

[4] B. Schouten, M. Tistarelli, C. Garcia-Mateo, F. Deravi, and M. Meints, "Nineteen urgent research topics in biometrics and identity management," in *Biometrics and Identity Management*, 2008, pp. 226–235.

[5] J. N. Pato and L. I. Millet, *Biometric Recognition: Challenges and Opportunities*. Washington, D.C.: The National Academies Press, 2010.

[6] E. P. Kukula, M. J. Sutton, and S. J. Elliott, "The Human–Biometric-Sensor Interaction Evaluation Method: Biometric Performance and Usability Measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 784–791, Apr. 2010.

[7] S. K. Modi, S. J. Elliott, and H. Kim, "Statistical analysis of fingerprint sensor interoperability performance," in *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1–6.

[8] D. Impedovo and G. Pirlo, “Automatic signature verification: the state of the art,” *IEEE Transactions on Systems, Man, and Cybernetics, …*, vol. 38, no. 5, pp. 609–635, 2008.

[9] K. S. Wilson and T. R. Callaghan, “Handheld Computer Terminals: Starting off Right the First Time,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1994, vol. 38, no. 5, pp. 355–359.

[10] E. P. Kukula, S. J. Elliott, and V. G. Duffy, “The effects of human interaction on biometric system performance,” in *First International Conference on Digital Human Modeling (ICDHM 2007), Held as Part of HCI International*, 2007, pp. 904–914.

[11] M. Brockly, R. Guest, S. Elliott, and J. Scott, “Dynamic Signature Verification and the Human Biometric Sensor Interaction Model,” *45th annual IEEE International Carnahan Conference on Security Technology*, pp. 253–258, 2011.

[12] A. Wamsley, S. Elliott, C. Dunkelberger, and M. Mershon, “Analysis of slap segmentation and HBSI errors across different force levels,” in *2011 Carnahan Conference on Security Technology*, 2011, pp. 1–5.

[13] S. J. Elliott, N. G. Street, W. Lafayette, E. P. Kukula, and J. M. Werner, “An Evaluation of the Human Biometric Sensor Interaction using Hand Geometry,” 2010.

[14] S. Elliott, E. Kukula, and N. Sickler, “The challenges of the environment and the human/biometric device interaction on biometric system performance,” *… forum on Modeling and Simulation in …*, 2004.

[15] E. Kukula and S. Elliott, “Implementing ergonomic principles in a biometric system: a look at the human biometric sensor interaction (HBSI),” *Carnahan Conferences Security …*, pp. 86–91, 2006.

[16] International Organization for Standardization (ISO)., “Information technology – Biometric performance testing and reporting – Part 6: Testing methodologies for operational evaluation. ISO/IEC FCD 19795-6.2,” 2011.

[17] “Biometric Performance Testing and Reporting — Part 1: Principles and Framework. ISO/IEC FCD 19795-1,” 2005.

[18] R. Micheals, B. Stanton, M. Theofanos, and S. Orandi, “A Taxonomy of Definitions for Usability Studies in Biometrics.” NIST, Gaithersburg, p. 9, 2006.

[19] J. W. M. Campbell and M. J. Madden, “ILO Seafarers’ Identity Documents Biometric Interoperability Test (ISBIT-4) Report,” 2009.

[20] M. E. Brockly and S. J. Elliott, “Automatic Detection of Biometric Transaction Times,” in *The 8th International Conference on Information Technology and Applications (ICITA 2013)*, 2013, no. Icita.

[21] L.R. Sudha and R. Bhavani "A combined classifier kNN-SVM in gait-based biometric authentication system", Int. J. of Computer Applications in Technology, 2014 Vol.49, No.2, pp.113 - 121

[22] Q. Mao, Y. Zhan, "Extraction and analysis of the speech emotion features based on multi-fractal spectrum", Int. J. of Computer Applications in Technology, 2010 Vol.38, No.1/2/3, pp.19 - 26