

Development of a Biometric Testing Protocol for Dynamic Signature Verification

Stephen J. Elliott

Abstract-- This paper accounts the testing protocols used at the author's university. It examines the experiences of one particular study in dynamic signature verification. The paper also outlines some additions to the current UK Biometric Working Group Best Practice document.

Index Terms—Curriculum, Undergraduate Education, Laboratory Activities, Graduate Education, Testing and Evaluation.

I. INTRODUCTION

THIS paper will outline the process and procedure for developing a dynamic signature verification testing protocol. The study has now been concluded; the results of the study can be found in Elliott (2002), but the testing protocol that has been designed is now used as a template for other biometric testing at the author's university. Additionally, the development of the testing protocol poses some additional factors to be addressed in the development of the protocol, as well as information required as part of the human subjects approval process within the university. This paper which is intended to be used as a discussion on testing protocol, is divided into five sections: a definition of terms and introduction to hypothesis testing, a discussion on the research questions, issues surrounding dynamic signature verification devices, and an evaluation of the pilot test protocol.

II. DEFINITION OF TERMS

1) General Biometric Performance Parameters

When evaluating and testing biometric devices, the measurement of five parameters are typically proposed:- the false match rate, false non-match rate, binning error rate, penetration coefficient rate and transaction times. Moreover, it is useful to have some measurements on the "failure to enroll" and "failure to acquire" rates. In testing devices, it is important to test them with the target application in mind.

According to Dunn (1998), false match rate is the percentage of impostors wrongly matched on a single comparison. False non-match rate is the percentage of valid users wrongly rejected. Equal error rate is when the false match rate equals the false non-match rate. Transaction rate is the amount of time required to complete the transaction. Failure to acquire is the rate to which the device has failed to

acquire a sample, the failure to enroll is the rate to which the device has failed to enroll an individual. The variables and their relationship among each other have consequences for determining the appropriate technology for specific applications.

2) Negative and Positive Identification

There are two applications of biometric technologies: positive identification, which proves you are who you say you are, and negative identification, which proves you are not who you say you are not. For positive identification, verifying the claim of the individual is through the comparison of the sample to an enrolled template. In a negative identification system, the user makes no claim to identity, therefore requiring a search of the entire database (UKBWG, 2000). Therefore, when enrolling in a negative identification system, there is a comparison of the enrollment template with all other enrollment templates in the system to make sure that there is not a match (UKBWG, 2000).

Positive identification does not require the use of biometrics. Other forms of physical identification, such as drivers' licenses, passports, and passwords, can positively identify the individual. Conversely, negative identification can only be achieved using biometrics (Wayman, 1999). Some applications require the use of negative identification, such as biometrics on commercial drivers' licenses. The one-driver, one-license, one-record goal requires a form of negative identification. When enrollment occurs in a negative identification system, the system compares the samples with all the templates in the database to ensure that there are no duplicate records (Wayman, 1999).

3) Hypothesis Testing

Statisticians use hypothesis testing to test "two formulations to be made on objective terms, with a knowledge of the risks associated with reaching the wrong conclusion" (Montgomery, 1996).

Two kinds of errors are made when using hypothesis testing; the first is when the null hypothesis is rejected when it is true, defined as a Type I error. The second error is not rejecting the null hypothesis when it is false, defined as a Type II error. The relationship between negative and positive identification and Type I and Type II errors relate to the determination of the hypotheses. Under positive identification (I am who I say I am), a Type I error occurs when rejecting the null hypothesis when it is true. For negative identification, the reverse is true, as shown in Table 1.

Manuscript received May 20, 2002

S. J. Elliott is with the Department of Industrial Technology, Purdue University, West Lafayette, 47907 USA (telephone: 765-494-1101, e-mail: sjelliott@tech.purdue.edu).

Table 1
Positive and Negative Identification and Type I and Type II Errors

	Positive Identification	Negative Identification
Type I	False Non Match	False Match
Type II	False Match	False Non Match

III. RESEARCH QUESTIONS

The development of a set of hypotheses centers on a specific idea. To illustrate this logical progression through the research process, several questions were asked of the researcher. What was the specific question of the study? Where is the data originating, what clearance, or process would be required to gain access to the population? What was the enrollment selection? Were there any issues with dynamic signature verification testing previously highlighted in the research?

1) *General Research Question*

The focus of the study was to examine the differences in the individual variables of the signature when signed dynamically across mobile computing devices. From this initial question, many others were formulated in order to develop a robust testing protocol. The first is device selection. Various methods of data acquisition are available to the researcher, outlined in Dullink, van Daalen, Nijhuis, Spaanenburg and Zuidhof (1995); Elliott, (2002); Greiss (2000); Han, Chang, Hsu, and Jeng, (1999); Hamilton, Whelan, McLaren, and MacIntyre, (1995); Holmes (1991); Komiya and Matsumoto (1999); Leclerc and Plamondon (1994); Lee, Berger and Aviczer (1996); Li, Parizeau and Plamondon (1998); Martens and Claesen (1997); Mingming and Wijesoma (2000); Mohankrishnan, Lee and Paulik (1998); Nanavati and Radke (2000); Narayanaswamy, Hu, Kashi, (1998); Plamondon (1994); Schmidt and Kraiss (1997); Wu, Jou and Lee (1997); and Yamazaki, Mizutani, and Komatsu (1998).

IV. ISSUES DIRECTLY RELATING TO DEVELOPING THE TESTING PROTOCOL

Further questions that arise in dynamic signature verification include the context of the signature (for example at a retail location or when signing a mortgage application), the location of the signing, habituation of the user to the signature device and the act of signing on that specific device.

Context of Signing

For example, the signature applied to a credit card slip at the supermarket may not be given the same amount of "thought" as that of a mortgage application, although in some cases the law does not provide for the electronic signing of a mortgage or will due to the required ceremony (personal communication, August 22, 2000). The National Physical Laboratory also conducted a review of a signature verification testing protocol; an excerpt follows.

In terms of the test protocol, we note a series of signatures given in quick succession will be very similar when compared to signatures separated by hours or days. We might ask whether the repeated attempts really give much more independent data for analysis. I note also that some signature systems recommend basing the template on signatures collected on different days (personal communication, 2001).

There are other factors to consider when developing a testing protocol for dynamic signature verification. Depending on the application and specific biometric technology, other questions may also arise. Although specific to the dynamic signature verification study, asking these questions will enable the researcher to modify the testing protocol and are transferable to other technologies. Such questions include asking that the user walk around (to change his/her position) between signatures, instructing to sign slowly on some occasions and faster on others, and collecting signatures on different pieces of paper (possibly with different properties, e.g. roughness, slipperiness, size) (personal communication, 2000).

Environmental conditions are also important to consider when designing a testing protocol. Examples include the environment of the test: What height is the counter? Is it angled or horizontal? What is the size of the signature box on the paper? Temperature - it is harder to sign with cold hands. These questions should also be asked for when developing a testing protocol for other biometric technologies.

Habituation

Discussion on user habituation to the device is important. People typically enroll their signature without being familiar to the device and/or the software. According to Mettyear, (personal communication, June 13, 2000), there is considerable long-term variation in signatures as well as a difference between morning and evening signatures; and there are hardware differences. Mettyear (2000) states that:

Problems can be caused by data obtained from participants who are not familiar with the hardware being used. Digitizers vary greatly in geometry and in the quality of signing surface; the sensory feedback received is quite different from signing on paper and on some devices there are buttons and other features that the hand must avoid so that is difficult to sign naturally. This can be off-putting for the novice user and there is evidence that some individuals have to sign many times over a period of weeks before they feel comfortable with new equipment. During this period, their signature may not be considered to be truly representative and should not be used to enroll a template (personal communication, July, 2000).

Anonymity

The Best Practice Document from the National Physical Laboratory in the United Kingdom is an excellent guide to the development of testing protocol. There are instances when the document does not apply to the specific nature of your

problem. All deviations from this protocol should be noted. An example of a limitation to the best practice document is illustrated with the example of dynamic signature verification and the classification of the impostor.

One such limitation with the Best Practice document is Section 32 (UKBWG, 2000). Some aspects of the software that are available (forensic tools) reconstitute the signature from the dynamic data so that the signature can be checked to make sure that the correct signature was signed. Without the use of this checking ability, the investigator would not be able to screen any of the samples. However, in an experiment using a limited population, signing of the name reveals the identity of the individuals (assuming that they were signing their own name). While the requirement of the section is that the "identities of the crew are never released" (p.6) those involved in signature studies will have to make sure that their procedures and protocols allow for the security of the signature, and its raw data. The addition of a protocol for dealing with anonymity into the Best Practice document would enable human subjects committees within a university environment to be reassured over the storage and security of this data.

Metyear (2000) suggests that there are levels of information that the signer might have in order to make an attempt at a forgery. He proposes seven different levels shown in Table 2.

Table 2
Impostor Knowledge About the Signature Data

Level	Information Available
0	B has no relevant knowledge of A
1	B knows A's name
2	B has seen a static image of A's signature prior to signing
3	B can see a static image of A's signature at the time of signing
4	B is able to trace a sample of A's signature
5	B has recently witnessed A's signing
6	B has repeatedly witnessed A signing

Table 2 indicates the relationship between the prior knowledge of the signature and the success of the forgery. Without this information, it is impossible to assess the meaning of the Type I and Type II errors shown in Table 3. Data from Table 3 comes from a summary of work in Plamondon (1994). Assuming the studies are measuring positive identification applications, the Type I and Type II errors are all measuring the same hypotheses. However, there is no indication on the proficiency of the impostor.

Further examination of the individual studies show that it is important to detail the knowledge of the impostor.

Komiya and Matsumoto (1999) had a database consisting of 293 genuine writings and 540 forgery writings from eight individuals. The study did not indicate how the forgery took place, except that they used the same eight individuals. Schmidt and Kraiss' (1997) study was comprised of 496 original signatures from 27 people. Each person signed 11 to 20 times. The database contained 48 forgeries that "fulfill the requirement on the visual agreement and the dynamic

similarity with the original signature" (Schmidt and Kraiss, 1997. p.5). Lee, Mohankrishnan and Paulik (1998), trained the algorithm using 250 signatures per writer, 100 were authentic signatures, and 150 were random forgeries classified as the genuine signatures of other writers. Wu, Jou and Lee (1997) used 27 people in the study, writing their own signature. The study also used four people imitating the signatures of all registered people. However, no further information was given on the selection of the impostor, or what information they were given in order to forge the signature. Hamilton, Whelan, McLaren, and MacIntyre (1995) used real signatures from other individuals as forgeries. In addition, a group of synthesized signatures was created by distorting real signatures through the addition of low-level noise and dilation/erosion of the various structures of the signature. Mingming and Wijesoma (2000) motivated the forgers by giving a cash reward. Han, Chang, Hsu and Jeng (1999) examined people's signatures over a four-month period.

Table 3
DSV Studies and Error Rates

Authors	Error Rates		
	Type I (FAR)	Type II (FRR)	EER
Achemlal, Mourier, Lorette, Bonnefoy (1986)	11.0%	8.0%	
Beatson (1986)	1.0%	2.0%	
Bechet (1984)	5.0%	5.0%	
Bonnefoy, Lorette (1981)	0-6%		
Bault, Plamondon (1981)	1.2%	1.0%	
Collantier (1984)	3.5%		
Crane, Ostrem (1983)	1.5%	1.5%	
Debruyne (1985)	3.0%	2.0%	
Hale, Pagnini (1980)	1.5%	1.2/2.5 %	
Herbst, Liu (1979)	1.7%	0.4%	
Herbst, Liu (1979)	2.4%	3.4%	
Ibrahim, Levrat (1979)	19.0%	5.5%	
Lam, Kamis (1989)	0.0%	2.5%	
Lorette (1983)	6.0%		
Mital, Hin, Leng (1989)	0.0%	0.0%	
Parizeau, Plamondon (1989)	4.0%		
Sato, Kogure (1982)	1.8%	0.0%	
Worthington, Chainer, Williford, Gundersen (1985)	1.8%	0.28-2.33%	
Zimmerman, Varady (1985)	30-50%	4-12%	
Cordella, Foggia, Sanson, Vento	0.03-20.82%	5.7%	
Wirtz (1997)			10%
Dimauro, Impedovo, Pirlo (1993)	1.7%	1.2%	
Nalwa (1997)			3%
Nalwa (1997)			2%
Nalwa (1997)			5%
Mingming, Wijesoma (2000)			5%
Mingming, Wijesoma (2000)			7%

Mingming, Wijesoma (2000)			9%
Hamilton, Whelan, McLaren, MacIntyre (1995)	34.0%	26.0%	
Hamilton, Whelan, McLaren, MacIntyre (1995)	22.0%	18.0%	
Hamilton, Whelan, McLaren, MacIntyre (1995)	12.0%	10.0%	
Hamilton, Whelan, McLaren, MacIntyre (1995)	7.0%	6.0%	
Martens, Clausen (1997)	1.5%	1.3%	
Chang, Wang, Suen	2.0%	2.5%	
Higashino	8.0%	0.6%	
Minot, Gentric	2.0%	4.0%	
Desjardin, Doux, Milgram	0.0%	0.4%	
Lucas, Damper	5.6%	4.5%	
Tseng, Huang	12.5-28.8%	5.0-12.5%	
Lee, Berger, Aviczer (1996)	1.0%	20.0%	
Lee, Berger, Aviczer (1996)	5.0%	20.0%	
Lee, Mohankrshnan, Paulik (1998)	0.9%	0.7%	
Han, Chang, Hsu, Jeng (1999)	4.0%	7.2%	
Komiya, Matsumoto (1999)			2%
Cardot, Revenu, Victorri, Revillet (1993)	2.0%	4.0%	
Cardot, Revenu, Victorri, Revillet (1993)	0.9%	7.4%	

V. TEST PROCEDURES

This section shows the testing procedures for the dynamic signature verification study, outlining the data collection design, the volunteer crew recruitment, and membership. Many of these subsections also satisfy some of the requirements for human subjects approval required by the author's university.

1. Data Collection Procedure Design

Volunteer Crew

The crew membership (subjects) needs to be defined; - the example used is from the dynamic signature verification study (Elliott, 2002). Crew membership was limited to those over the age of 18, the minimum age to hold a credit card. Upper age limits and sex were not important in the membership of the crew. The study, conducted in the School of Technology's Department of Industrial Technology, has a higher percentage of white males than minority males, white females, or minority females. Consequently, there was a higher rate of white males among the volunteers.

Recruitment of Subjects

Recruitment of the subjects needs to be noted, for example were they self selecting, or paid? Many studies do not include the recruitment methodology for the volunteer crew, however, as shown in the study of dynamic signature verification studies it is interesting to note so that replication of the study can be achieved, as well as other variables

Confidentiality

A statement of confidentiality and consent form, outline of the study, and the procedures for the subject should be noted. Subdivisions addressing potential risks to the subjects, gains by the individual and/or society, and the investigators' evaluation of the risk-benefit ratio are also noted.

VI. THREATS TO INTERNAL AND EXTERNAL VALIDITY

Internal validity is the assertion that the observed effect is due to the independent variable(s) in the study. External validity is the generalizability of the study's findings to other populations, places, or times (McMillan and Schumacher, 1997). The experimental design has threats to both internal and external validity, and a tradeoff occurred.

History refers to "extraneous incidents or events affecting the results that occur during the research" (McMillan and Schumacher, 1997, p.184). Within the context of this study, historical factors did not influence the study. Members of the crew that did know about biometrics from media outlets were not influenced by them in such a way to affect the outcome of the study. Selection is the second factor influencing the design of the study; the crew was self-selecting which was unavoidable due to the requirements of the Purdue University Human Subjects Committee. Instrumentation concerns the "way the changes in the instruments or persons used to collect data might affect the results" (McMillan and Schumacher, 1997, p.186). All devices and measurements remained constant, reducing the effects of the instrumentation effect. Subject attrition or mortality effects occur when there is attrition from the group. The college calendar works on a 16-week cycle; therefore, to reduce the amount of attrition, the study ended within the semester (16 week) period. Additionally, volunteers were more likely to miss a meeting later on in the semester, as other work such as projects, mid-term exams, and meetings take priority for the volunteer crew. Experimenter effects refer to "both the deliberate and unintentional influence that the researchers has on the subjects" (McMillan and Schumacher, 1997, p.188). Additional internal validity factors influencing the design of the study included treatment replications and subject effects. Treatment replications occurred over a period of one to 45 days. To ensure equal treatment replications, the study used the same testing protocol.

External validity threats include population and ecological validity. Population validity is the extent to which the results can be generalized only to other people who have the same, or at least similar, characteristics as those used in the experiment" (McMillan and Schumacher, 1997, p.190). Ecological external validity "refers to the conditions of the research and the extent to which generalizing the results is limited to similar conditions" (McMillan and Schumacher, 1997, p.190). Chapter 5 describes the environment and the volunteer crew. The construction of the experiment adequately weighed both threats to internal and external validity.

VII. PILOT STUDY

A pilot study is a useful tool to recommend changes to the testing protocol before the full scale testing. It enables the researcher to test the software, the environmental conditions, and gain subject feedback. Some examples of the results from the pilot study include the following.

- Signing over 100 times on the mobile devices to test the data holding capacity of the devices
- Signing on the e-pad digitizer and Wacom digitizer to test for errors
- Timing the volunteers in order to calculate the throughput time during the study
- Testing the layout of the experimental area, for both right and left handed volunteers
- Implement an additional reference signature at signing to test for degradation of variables
- Move the monitors that display real-time information on signing so that the original device requirements were met. One subject remarked that (s)he altered their signature so that it would "appear" normal.

Room layout recommendations included the following

- Reconfigured room layout accommodating left handed subjects
- Fixing one portable digitizer to the table. If an individual moved the device, repositioning occurred according to the template plan shown below in the room preparation section
- Code 39 bar codes printed on the back of the Human Subject forms resulting in a reduced possibility of incorrect keying on the part of the researcher

Room specifications were examined, and the layout of the experimental area changed between the pilot study and the actual data collection because of a number of observations from the volunteers, and the lack of space.

VIII. CONCLUSION

Designing a testing protocol is an important part of the testing strategy. The paper outlines the development of a testing protocol used in a dynamic signature verification study (Elliott, 2002).

REFERENCES

- Dullink, H., van Daalen, B., Nijhuis, J., Spaanenburg, L. & Zuidhof, H. (1995, December). Implementing a DSP Kernel for Online Dynamic Handwritten Signature Verification Using the TMS320DSP Family (SPRA304, pp. 1-26). France: EFRIE.
- Elliott, S. J. (2002). A Comparison of On-Line Dynamic Signature Trait Variables vis-a-vis Mobile Computing Devices and Table-Based Digitizers. Auto ID 2002: Workshop on Automatic Identification Advanced Technologies (pp.121-125), Tarrytown, NY
- Dunn, J. (1998, May 20). Hearing on Biometrics and the Future of Money. In Committee on Banking and Financial Services. Washington DC: House of Representatives. Retrieved November 24, 2000 from the World Wide Web:
<http://www.house.gov/banking/52098jd.htm>
- Greiss, F. D. (2000, May). On-Line Signature Verification. East Lansing, MI: Michigan State University, Department of Computer Science.
- Han, Chang, Hsu, and Jeng, (1999). An On-Line Signature Verification System Using Multi-template Matching Approaches, submit to IEEE International Carnahan Conference on Security Technology, 1999. Madrid, Spain.
- Hamilton, D. J., Whelan, J., McLaren, A. & MacIntyre, I. (1995). Low Cost Dynamic Signature Verification System. In European Convention on Security and Detention (pp. 202-206). London, UK: IEE.
- Holmes, J. P. (1991). A Performance Evaluation of Biometric Identification Devices. Scandia National Labs. (SANDIA91-0276)
- Komiya, Y. & Matsumoto, T. (1999). On-line pen input signature verification PPI (pen-position / pen-pressure / pen-inclination. IEEE. in, C., Barcelo, R., Baker, S. & Greenwald, E. (2000, Sept). An Analysis of International Electronic and Digital Signature Implementation Initiatives. In Internet Law and Policy Forum. New York, NY: Internet Law and Policy Forum. Retrieved November 24, 2000 from the World Wide Web:
http://www.ilpf.org/digsig/analysis_IEDSII.htm
- Leclerc, F. & Plamondon, R. (1994). Automatic Signature Verification: The State of the Art - 1989-1993. *Progress in Automatic Signature Verification* (pp. 3-21). Singapore: World Scientific Publishing Co.
- Lee, L., Berger, T. & Aviczer, E. (1996). Reliable On-Line Human Signature Verification Systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 643-647.
- Li, X., Parizeau, M. & Plamondon, R. (1998). Segmentation and Reconstruction of On-Line Handwritten Scr. *Pattern Recognition*, 675-684.
- Martens, R. & Claesen, L. (1997). On-Line Signature Verification: Discrimination Emphasised. IEEE.
- McMillan, J. & Schumacher, S. (1997). Research in Education: A Conceptual Introduction (4th Ed), Addison-Wesley, New York, NY
- Mettyear, N. (2000). Error Rates in Biometric User Authentication. Unpublished manuscript
- Mingming, M. & Wijesoma, W. (2000). Automatic On-Line Signature Verification Based on Multiple Models. In CIFE'01: Computational Intelligence in Financial Engineering Conference (pp. 30-33).IEEE.
- Mohankrishnan, N., Lee, W.-S. & Paulik, M. (1998). Improved Segmentation through Dynamic Time Warping For Signature Verification using a Neural Network Classifier. In IEEE Signal Processing Society International Conference on Image Processing. IEEE Signal Processing Society. (Original work published 1998)

- Montgomery, D. (1996). Design and Analysis of Experiments. 5th Edition, John Wiley and Sons.
- Nanavati, S. & Radke, M. (2000, April). Dynamic Thresholding. In Research Paper. New York, NY: International Biometric Group. Retrieved March 24, 2000 from the World Wide Web: <http://www.ibg.com/>
- Narayanaswamy, S., Hu, J. & Kashi, R. (1998). User Interface for a PCS Smart Phone. In Proceedings of the IEEE International Conference on Multimedia Computing and Systems Volume I.IEEE.
- Plamondon, R. (1994). The Design of an On-Line Signature Verification System: From Theory to Practice. Progress in Automatic Signature Verification. New York: World Scientific Publishing.
- Schmidt, C. & Kraiss, K. F. (1997). Establishment of Personalized Templates for Automatic Signature Verification. IEEE, 263-267.
- UKBWG. (2000). United Kingdom Biometric Best Practice Document. UK Biometric Working Group, CESG. Retrieved November 23, 2000 from the World Wide Web: <http://www.afb.org.uk>
- Wayman, J. L. (1999). Fundamentals of Biometric Authentication Technologies. In J. L. Wayman (Ed.), National Biometric Test Center Collected Works (Vol. 1). San Jose, CA: National Biometric Test Center.
- Wu, Q.-Z., Jou, I.-C. & Lee, S.-Y. (1997). On-Line Signature Verification using LPC Cepstrum and Neural Networks. IEEE Transactions on Systems, Man, and Cybernetics, 148-153.
- Yamazaki, Y., Mizutani, Y. & Komatsu, N. (1998). Extraction of Personal Features from Stroke Shape, Writing Pressure and Pen Inclination in Ordinary Characters. Proceedings of the Fifth International Conference on Document Analysis and Recognition.