N. M. Herbst
C. N. Liu

Automatic Signature Verification Based on Accelerometry

Abstract: The fine structure of the muscle forces that are exerted during the writing of a signature is constant and well defined for most people. In general, the fine structure is not subject to conscious control. Based on these observations, an experimental system has been designed that utilizes a person's signature dynamics to verify identities. The design and operational features of this system are described. Experiments on 70 subjects during a four-week period show a 2.9 percent rejection of valid signatures and a 2.1 percent acceptance of forgeries. An average of 1.2 trials was necessary for verification. The forgers were knowledgeable about the verification technique and did their best to deceive the system. The acceptance rate of random forgeries, i.e., accidental matching of two separate signatures, was 0.16 percent.

Introduction

An effective automatic method of online signature verification will have many important applications. A method of personal identification that cannot be lost, stolen, or forgotten is required for control of computer access, building access or automatic banking. Because the signature is the normal and customary way of identifying an individual in our society, it has many natural advantages over competing techniques such as fingerprints or voice verification.

Document examiners have long realized that signatures, more than other kinds of writing, are written from habit [1, 2]. The writer is generally thinking about what he is signing rather than how to spell his name or form the characters. From a very early age, shortly after learning to write, children personalize their signatures, i.e., vary them from copybook style. Signatures are written for identification rather than legibility.

On the other hand, the track of the pen shows a great deal of variability. No two genuine signatures are ever precisely the same. Two identical signatures constitute legal evidence of forgery by tracing. The normal variability of signatures constitutes the greatest obstacle to be met in achieving automatic verification.

Signatures vary in their complexity, duration, and vulnerability to forgery. Signers vary in their coordination and consistency. Thus, the security of the system varies from user to user. A short, common name is no doubt easier to forge than a long, carefully written name, no matter what technique is employed. Therefore, the system must be capable of "degrading" gracefully when supplied with inconsistent signatures, and the security risks must be kept to acceptable levels.

This paper describes an online signature verification system based on acceleration measurements. The qualitative considerations underlying the relevance of accelerometry, which results from the nature of the muscular activity, are first described in the second section, and the verification system is presented in the third section.

For automatic comparison, two signatures must be registered in time, and certain gross distortions of the time axis must be removed. Furthermore, we must define a function that expresses the proper similarities quantitatively, e.g., high for valid signatures and low for forgeries. Our solution, regional correlation, is based on our model.

A practical verification system must also be able to set its parameters on the basis of a very small set of known signature samples. A technique based on the principle of optimizing the verification performance on the reference samples is described in the section entitled "Reference design procedures." The results of a reasonably large laboratory experiment evaluating these techniques are included. The results show, subject to the usual caveats of sample size and extensibility, that if one can overcome certain systematic failures, performance and human acceptability will be high.

Previous workers have made dynamic measurements on the signature process by using the conventional pattern recognition methodology of statistical decision. However, these measurements were unreliable and achieved only indifferent results. (We mention only online systems. Verification by optically scanning existing signatures has been even less successful.) Mauceri [3] took 50 signatures from each of 40 subjects, used power

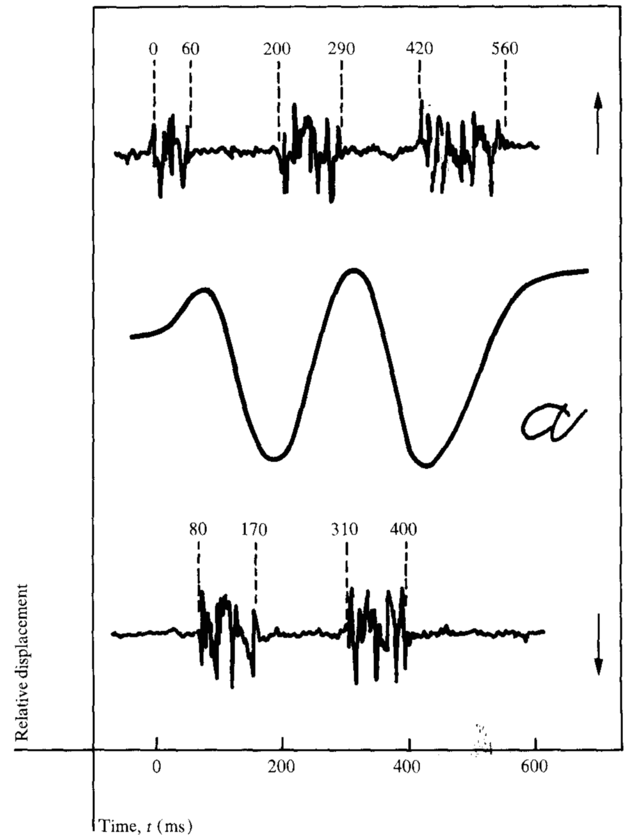

Figure 1 Displacement time diagram for movement perpendicular to direction of motion for the handwritten letter "a," together with associated electromyograms, after Vredenbregt and Koster [14]. Alternate nature of antagonist pair activity is evident.

spectral density and zero-crossing features, and was able to correctly classify a signature 63 percent of the time. Farag and Chien [4] used chain-encoded tablet data as input to a recognition scheme. The performance for ten subjects was 27 percent reject, 27 percent success of forgery.

More recently, a system based on pressure measurements was announced [5]. The prototype is now being tested by the U.S. Air Force. Performance, based on three signatures per trial, is given as 6.81 percent reject and 3.19 percent random forgery [6]. A "random forgery" consists in trying every valid signature in the data base against every other, regardless of signer. It substantially underestimates the actual forgery rate. When actual forgery was tried, 10 out of 58 attempts were successful, i.e., 17 percent. In the same tests [6], an automatic speaker verification method was evaluated and gave better results (2.5 percent reject and 0.64 percent random forgery, operating adaptively). The training procedure is relatively involved, requiring the utterance of a minimum of twenty nonrepetitive four-word phrases. Although subsequent verification is rapid and leaves the hands free, the procedure requires special preliminary training sessions.

Although there are no definitive experiments, human examination of signatures is not very accurate. Trained document examiners were reported in one test as achieving 25 percent no-opinion or reject, without accepting any forgeries [7]. Untrained personnel such as bank tellers accept 10–50 percent of the forged signatures presented when under test conditions [5], and almost 100 percent in actual practice. Most of the successful forgeries that occur are totally unskilled. Bank losses through forgery of stolen checks were estimated by the American Bankers Association at 50 million dollars for 1974, far exceeding the total losses due to bank robbery and burglary combined [8].

Models of handwriting

Motions controlled by sensory feedback are generally slow and precise. Both opposing muscles (called the agonist and the antagonist) for a particular degree of freedom are active together, and their ratio is controlled consciously. Since the muscles are organized in groups, accurate measurement of these motions shows stepwise increases in force rather than a smooth continuous motion. Other concomitants of closed-loop control systems, such as hunting (tremor) and instability (in neurological diseases), can also be observed. Much writing and drawing clearly consists of controlled motions.

But not all bodily motions are controlled by sensory feedback. Those motions that do not involve sensory feedback are called ballistic motions. These are generally rapid, practiced motions whose accuracy increases with speed [9]. In a sense, they cannot be done slowly at all. Walking, playing a musical instrument, and tennis or golf swings are all examples of ballistic motions. In many cases, a motion can be done either consciously or ballistically, albeit at different speeds. The purest example of ballistic motion is the rapid saccadic motion of the eye [10]. The saccades, or small jumps, are typically 10–30 ms in duration.

The individual muscle forces in rapid handwriting are 30–100 ms in duration. Sensory feedback from the eye to the brain to the hand requires on the order of 200 ms. The individual muscle forces, therefore, cannot possibly be determined by simple feedback but are rather predetermined by the brain. This can also be demonstrated experimentally [11]. These forces are not only predetermined but are given strictly in terms of only two variables—magnitude and duration.

When applied to handwriting, the ballistic model gives a better fit to measured data than does any model yet proposed [11–13]. Vredenbregt and Koster [14] built a simple simulator based on the ballistic model, containing damping in the form of spring forces arising from the stiffness of the unexcited opposing muscle, and a viscous damping term representing the various fluids surrounding the muscles. The stiffness term is negligible and the viscous damping term is assumed to be constant although, as Yasuhara later pointed out [13], it contains secondary effects from friction, and hence from pressure. The excitations were programs of impulses deduced from the envelope of measured electromyograms, shown in Fig. 1. The simulator wrote single characters quite naturally. Most significantly, perturbations in the excitation gave rise to natural-looking distortions in the resulting pattern, and changes as short as five ms produced alterations in the character shape. This simulation suggests that the muscles are excited with impulses of considerable accuracy. These notions have not previously been applied to signatures.

Perhaps the most striking aspect of signature dynamics is that the time interval for writing a signature, measured from start to finish, remains remarkably consistent. Successive signatures frequently differ in duration by as little as 10 ms from each other. Combined with the assumption that signatures are ballistic motions, which implies that the motions are completely predetermined, this observation leads us to expect that the durations of the individual muscle forces are also identical. If the signature is indeed a constant-time phenomenon, the magnitude of the forces parallel to the writing surface is then related only to the size of the resultant trace, i.e., the distance the pen point traverses as a result of a given force. We then expect that the durations and hence the zero-crossings of the pen acceleration would be invariant. (Additional information is contained in the relative amplitudes of the strokes.) The corresponding strokes of different signatures will thus be formed by forces of identical duration. Changes in size are interpreted as changes in force [11]. Our first inspection of acceleration waveforms, Fig. 2, showed that this description was qualitatively correct.

Based on this model, we propose an automatic signature verification system using the acceleration-time function as the principal measurement.

For a variety of reasons, this concept is not completely straightforward to implement or test. Because of gross variations in the signature, a program written to compare signatures automatically may not actually compare the proper segments. Secondly, the above arguments are true for individual muscle groups, but the motion of the pen is the product of several different muscle groups: pivoting between the thumb and forefinger, pivoting at the wrist, and pivoting at the elbow and shoulder. Each antagonist pair may have its own timing, and we know nothing about the synchronization of the various combinations. Finally, the timing correspondence is clearly not absolute. There are variations on a day-to-day basis, even for valid signatures, and the time axis is subject to many distortions. There may be major pauses and hesitations, minor tics and gaps, added or deleted strokes, and even misspellings.

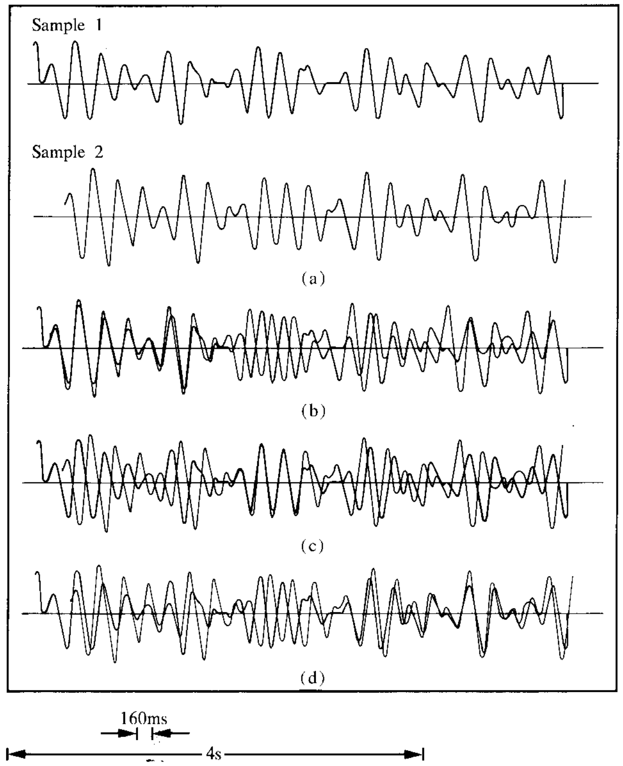

Figure 2 Accelerations in y-axis of plane of paper derived from tablet data for a) two samples of the same individual's signature, b) signals superimposed so that the first names correspond, c) signals shifted to match the middle section, and d) signals shifted to match the last section. Correspondence of force durations is striking. Amplitudes differ because two samples were of slightly different size.

Our decision technique was chosen to find the best time registration on a global basis by means of a modified correlation function. The initial registration was based on pen contact with the writing surface. Thus, we did not rely on the presence of any one particular landmark. Furthermore, correlation is sufficiently general to be evaluated on relatively small data sets. We were therefore able to have an operating verification system designed on a small sample while gathering a larger data base. Using correlation to find the registration does not preclude the use of a second stage to the decision process, based on a more complete analysis of the motion in terms of the foregoing model.

Verification system

A general signature verification system basically consists of four parts: the transducer, the comparator, the reference file, and the decision logic. It should be clear from the previous section that we wish to measure or derive the muscle forces in the hand during the signature. If a transducer of constant weight and shape is used, we may then observe the accelerations of the pen.

In the experiments reported here, an experimental pen containing two orthogonal piezoelectric accelerometers is mounted as close to the point as possible. It has a triangular cross section with a preferred orientation, and is tapered smoothly from the point to the top. The writing element is an ordinary ball-point cartridge. The writing surface is a paper tape one inch wide, fitted over a rectangular treadle that closes a set of gold contacts when a force of more than 30 g is applied. This serves as the pen switch. Making the pen switch separate from the pen itself simplified the pen structure for the purpose of this investigation. The signals are digitized at the rate of five ms, and read into an IBM System/7 computer. The signature is then transferred into a time shared IBM 370/145 under VM. The comparator is a PL/I correlation program which runs on the 370/145. The decision result is transferred back to the System/7 which lights a lamp to notify the signer. Under normal loads, the system responds in 10 s, which seems quite acceptable.

For the test, the reference library of signatures is kept on a separate disc file. References are usually generated offline in an overnight operation on the basis of a five-signature design set. Adaptation has not been permitted in these experiments, although it is definitely possible.

Signatures are 2 - 10 s in duration, with the average being about 5 s. The signals, sampled at the rate of 5 ms, 2 bytes/point, are thus 2000 bytes on the average. The waveforms are substantially oversampled in both dimensions. With run-length coding and a coarser amplitude grid, 100 - 400 bytes appear ample to describe the signature fully.

The dual accelerometer pen is sensitive at high frequencies, although it contains an electrical roll-off of 6 dB per octave at 30 Hz. A linear-phase recursive digital filter with integer coefficients has been designed to remove the high-frequency paper noise. It has a zero at 40 Hz and a half-power point of 22.5 Hz. The linear phase property maintains the timing of the zero-crossings undisturbed.

Regional correlation algorithm

To compare the sample and reference signatures, each is partitioned into pieces, called segments, and corresponding segments are cross-correlated with a modified correlation measure. Various segmentation policies are used. The average correlation depends on the segment length; the longer the segment, the more difficult it is to find alignment because of the accumulation of minor errors. Thus far, the best results on a large sample have been found with segments in the range of 1 - 2 s. This depends on the distribution of variable pauses in the signature; some individuals write entirely without pauses, and longer segments give satisfactory matches. Although the experiments are not yet conclusive, best results thus far have been obtained from segmentation based on paper contact. Variable pauses seem to occur more frequently when the pen is in the air. On the other hand, for many signatures the intervals while the pen is in the air contain characteristic motions that should be of discriminatory value since they are completely secret. Thus, a segment can be defined with equal validity to extend from pen-down to pen-down, including this time, or from pen-down to pen-up. This is a program variable.

Segmentation can also be accomplished without paper contact information. The signature may simply be equally divided (uniform segmentation), or various computational approaches can be taken to find matching segments (e.g., register the amplitude peaks, correlate continuously using a piece of the reference, etc.); the segments also can be allowed to overlap.

If the two orthogonal acceleration channels are labeled p and q, and the sample and reference signals are

$$s(t) = \langle s_p(t), s_q(t) \rangle \qquad \text{and } r(t) = \langle r_p(t), r_q(t) \rangle, \tag{1}$$

then for the ith segment, the correlation may be defined as

$$\begin{aligned} C^i &= \max_\tau \sum_t [s^i(t) r^i(t+\tau)] \\ &= \max_\tau \sum_t [s^i_p(t) + js^i_q(t+\tau)][r^i_p(t) - jr^i_q(t+\tau)] \\ &= |C| e^{j\phi}, \end{aligned} \tag{2}$$

where the segment is normalized such that

$$\sum_t s^i(t) = 0,$$

$$\sum_t r^i(t) = 0,$$

and the correlation is normalized by dividing it by

$$\left[\sum_t \left\{ [s^i_p(t)]^2 + [s^i_q(t)]^2 \right\} \sum_t \left\{ [r^i_p(t)]^2 + [r^i_q(t)]^2 \right\} \right]^{\frac{1}{2}}.$$

This is the complex correlation of magnitude C and phase angle ϕ, normalized so that the signals have zero mean and unit energy in the product. One advantage of this form of correlation is that its magnitude is rotationally invariant. Suppose the sample is rotated by θ, i.e., $s^* = se^{j\theta}$. Clearly, from the above equation, the magnitude of C is not changed from the unrotated value.

An obvious alternative is to correlate each channel individually and form an average for the total correlation. That is,

$$c_p^i(\tau) = \sum_t s_p^i(t) r_q^i(t+\tau), \qquad \text{and}$$

$$c_q^i(\tau) = \sum_t s_q^i(t) r_q^i(t+\tau), \tag{3}$$

with similar normalization.

Then we can with equal reason combine according to an arithmetic rule,

$$C^i = \tfrac{1}{2} \max_\tau \left(c_p^i(\tau) + c_q^i(\tau)\right), \tag{4}$$

or a geometric rule,

$$C^i = \max_\tau \sqrt{c_p^i(\tau)\ c_q^i(\tau)}. \tag{5}$$

These combinations reflect various degrees of stringency, because the geometric rule is less forgiving. Consider the forger who is perfect in one channel and zero in the other. The arithmetic formula will assign a correlation of 0.5 and the geometric will give zero.

The complex correlation automatically removes a fixed rotation of the pen, allowing the use of an unoriented pen. But this gain in freedom costs in terms of discrimination, since the forger would seem to have a somewhat less difficult task; hence, both methods must be evaluated experimentally.

The segments are cross-correlated in isolation, replacing the neighboring signals with padding zeros. The sample is shifted by up to 20 percent of the reference length, but never more than 300 ms. The correlation results are then weighted to penalize excessive shifting, because this improves discrimination [15].

The pen lift pattern, which may be viewed as the quantized axial pressure, contains a significant random component. While the principal pen lifts denoting different sections of the name should always be present, others can come and go. Although there are repetitive pressure patterns, document examiners have noted that writing pressure is varied consciously in response to how the pen is inking rather than from habit. Many individuals lighten up on retrace strokes for pictorial effect. Whether or not the pen inks on the retrace stroke is often not important to the signer. The presence of a pen lift will thus depend on the particular algorithm or hardware switch used.

The strategy we have implemented, called VPEN, is to eliminate extra pen lifts, trying all combinations exhaustively, until the best mean-square match between the two timing patterns is achieved; i.e., if the two signatures X and R have segment times T_X and T_R,

$$T_X = \langle T_1, T_2, \cdots, T_\mu \rangle;$$

$$T_R = \langle T_1, T_2, \cdots, T_\nu \rangle, \tag{6}$$

then we eliminate all combinations of pen lifts to make $\mu = \nu$, thus minimizing $|T_X - T_R|$, the root-mean-square norm. An exhaustive process is feasible because μ and ν are small, typically six or less.

The result of cross-correlating two signatures X and R is a vector consisting of a correlation for each segment,

$$C_{XR} = \langle C^1, C^2, \cdots, C^\nu \rangle, \tag{7}$$

where the C^i are the segment cross-correlations. Let n be the corresponding segment lengths of the reference signature. The results must then be combined to give the final decision. Properly speaking, we should weight each segment's cross-correlation according to the difficulty of forging it. This information is usually not available. However, we have noticed that forgers have more trouble with the longer segments. This seems natural because a longer segment represents a more complex motor task. Accordingly, each segment is weighted by the reference length.

We define V (for verification measure) as

$$V = \sum_i (n_r^i / n)\ C^i,$$

where $n = \Sigma_i n_r^i$ and n_r^i is the ith segment length in sample points for the reference signature. A threshold based on the analysis data must then be chosen.

Reference design procedures

Reference design may be viewed as a combinatorial optimization problem. Of the Comb (M, N) ways in which M signatures may be selected from N given signatures, the "best" M signatures must be determined as reference signatures for verification purposes. This statement implies that we need both a criterion by which each subset of M signatures can be evaluated and algorithms for determining the optimal subset as references. We describe in this section a selection criterion, an algorithm for reference design, and a procedure for estimating the decision threshold for each reference.

Let $s_1(t), s_2(t), \cdots, s_N(t)$ be N sample signatures given by a user. Our criterion is the following: We wish to find a minimal subset of M signatures such that with the signatures in this subset as reference signatures, the verification measure between each one of the remaining $N-M$ signatures and one of the M reference signatures will be at least as large as a specified value. In other words, we would like to maintain a certain specified value of verification measure between a set of reference signatures and the rest of the sample signatures. If more than one subset of M signatures satisfies this criterion, then the optimal subset is defined as the one in which the smallest verification measure between a sample and its nearest reference signature is the largest among all subsets of M satisfying the criterion.

Table 1 Comparison of correlation algorithms.

	Average of two correlations	*Magnitude of complex correlation*
Rejection of valid signatures	183/695 = 26%	124/695 = 17.8%
Acceptance of forged signatures	2/287 = 0.70%	2/287 = 0.70%

Table 2 Effect of multiple references per user.

	Single reference per user	*Two or fewer references per user*
Rejection of valid signatures	142/695 = 20.4%	124/695 = 17.8%
Acceptance of forged signatures	4/287 = 1.4%	2/287 = 0.70%

Table 3 Verification results with three trials.

Rejection of valid signers	*Acceptance of forgers*	*Average number of trials*
17/592 = 2.87%	2/97 = 2.1%	695/592 = 1.17

The reference design algorithm selects an optimal reference set by sequentially determining whether, for $M = 1$, 2, a subset of M signatures satisfies the criterion. If no subset of two signatures meets the threshold requirement, then the subset closest to meeting the requirement is selected.

We have implemented this algorithm, and for practical reasons have chosen M to be less than or equal to two. If a specific number of reference signatures is desired, e.g., one, the algorithm can be used to select the "best" signature. However, the specified design threshold value in this case is not guaranteed.

Another plausible way to derive a reference signature is to take the sample average for each user. Since it is possible to have many variations in the design samples, it is often not feasible to average all samples into a single reference signature. The average of poorly correlated signals tends to zero. A sequential scheme that avoids this problem by averaging clusters of samples into reference signatures was implemented, but the above method was experimentally superior.

• *Verification threshold estimation*

In conventional pattern recognition applications, samples of all pattern classes are given for reference design and decision boundary determination. Many techniques have been developed for determining the optimal decision boundaries. However, for signature verification, forged signatures are not available for the reference design, and therefore a verification threshold must be chosen on the basis of the given valid signatures alone. The only information available on forgery is our previous general experience.

The particular estimation scheme used is as follows. Let the minimum of the set of verification measures between $R_i(t)$ and each sample assigned to $R_i(t)$ be V^*. Let the standard deviation of these verification measures be σ. Then, the verification threshold T_i for reference $R_i(t)$ is estimated as

$$T_i = \min\{0.6, \max[(V^* - k\sigma), 0.48]\}. \qquad (8)$$

In other words, we estimate T_i from the verification measures obtained between $R_i(t)$ and samples $s_j(t)$, and then set an upper limit of 0.6 and a lower limit of 0.48 on the estimated value. In a case for which there is only one sample or no sample signature assigned to reference $R_i(t)$, the threshold T_i is rather arbitrarily set at 0.5.

The parameter k selects the tradeoff between rejects and errors. Depending on the application, it could be varied dynamically. In these experiments, $k = 1.5$.

Experimental results

In this section, we present the results of our investigation of an experimental signature verification system.

In April 1975, we collected on our experimental system 1332 signatures from 70 volunteers. To simulate a real application environment, each individual was asked to sign five times on the system in the first session to provide the sample signatures needed for reference design. No special training or practice was provided. Based on the five samples, a set of references was designed for each user and stored in the reference file. In the subsequent daily sessions, each individual was given up to three trials to have his signature verified by the system. Subjects were motivated by the prospect of free coffee. For reference design, 350 signatures were collected (5×70 users). The total number of test signatures collected was 695. In addition to the valid signatures, we also collected 287 forged signatures. The "forger" was free to practice the target signature. In some instances, the forger watched how the valid signature was signed and attempted to duplicate the gross features of the motion. Some forgers attempted to exploit flaws in the experimental system. All forgers were urged to write rapidly and freely if they wished to succeed. Thus, these forgeries were not casual, but part of an attempt to exercise the verification scheme vigorously.

• *Decision rules*

To decide whether a signature should be verified or rejected, two decision rules are used sequentially. The first rule checks to see if the difference in time between the signature and a reference is within a tolerance of 20 percent. The signature is rejected if the time difference is greater than the tolerance. Otherwise the signature is passed to the second rule which checks whether the verification measure between this signature and the reference is higher than the verification threshold associated with the reference. A signature is verified if the verification threshold is exceeded for any one of the reference signatures. Otherwise it is rejected. Since the first rule involves very little computation, the two rules are used sequentially.

• *Comparison of correlation algorithms*

The performances of the two correlation algorithms described earlier, the magnitude of complex correlation, and the weighted average of correlations from two channels, were investigated experimentally. References were designed with $M \leq 2$. In the case of complex correlation, a set of 119 reference signatures was selected for the 70 users, and in the case of average correlation of two channels, a set of 118 reference signatures was selected. All of the 695 test signatures and 287 forged signatures were tested. The results are shown in Table 1.

It is apparent that the complex correlation algorithm gave superior performance and it was therefore chosen as our experimental system. The significantly large difference in the number of rejects was due to the problem of stylus rotation. Although there was a preferred orientation for our stylus, some users did not always conform to the convention.

• *Performance of single reference*

A set of single references for each user was designed. The desired design threshold value of $Th = 0.5$ could not be satisfied for approximately 70 percent of the users. The verification thresholds for this subset of users were set at 0.48. The results of all test and forged signatures are shown in Table 2. The restriction of one reference signature per user produced a significant degradation of performance.

• *Verification results*

The results of the online experiment using complex correlation, three trials for verification, and the set of 119 references are shown in Table 3. An examination of reject rate for each individual revealed that for exactly half of the users, all valid signatures were verified on the first trial and all forged signatures were rejected. More than half of the rejected valid signatures came from a subset of seven users, four of whom signed their names with motions that could not adequately be detected by our present stylus (the accelerometer outputs were largely zero). When a user did need a second trial, his success rate for the second trial on the average was roughly 70 percent. When he went to a third trial, the average rate of success dropped to below 40 percent. This is another indication of a systematic mode of failure. A plot of the distribution of rejects for valid users is shown in Fig. 3. The gross averages are thus distorted by the small number of users who fail badly.

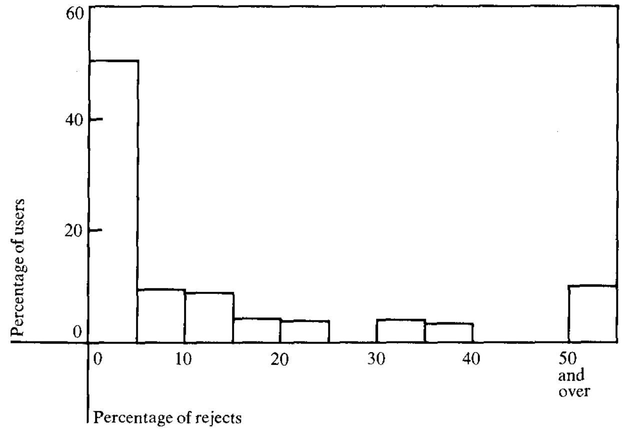

Figure 3 Distribution of individual reject rates. Over half the subjects never had a reject.

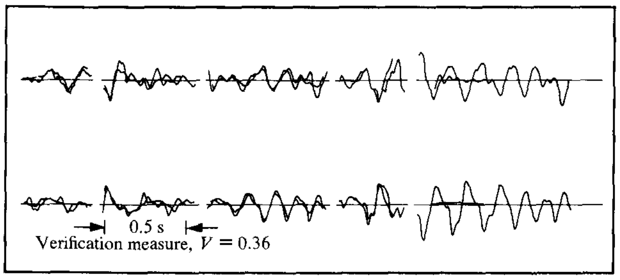

Figure 4 Typical rejected sample. Two signatures are shown overlapped in position of highest correlation. Each axis shows one channel. Rejection is caused by failure of individual to write his last name completely. Accelerometer outputs are shown for p channel, upper axis, and q channel, lower axis.

Figure 4 shows a rejected signature compared with its reference signature. Figure 5 shows an accepted forgery.

Another measurement of false acceptance rate was made by comparing the reference signatures of every user against the reference signatures of every other user. This is the random forgery experiment.

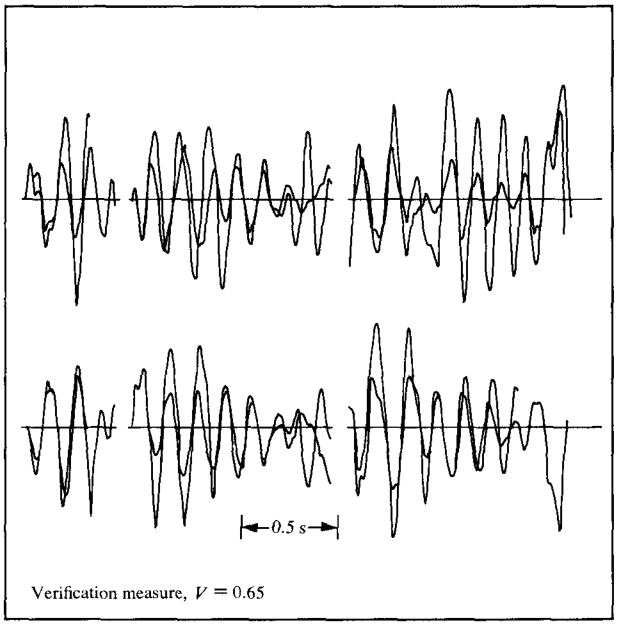

Figure 5 One of two accepted forgeries. Although timings match well, signals are still substantially different. The p channel is upper axis and q channel, lower axis.

The complete set of 119 reference signatures for 70 users was used in the simulation. Since each reference signature had an individual verification threshold, the total number of comparisons was 8211. It was found that there were 13 errors involving seven users' references, or an error rate of 0.16 percent. In 59 percent of the cases, forger rejections were based upon time discrepancies.

An examination of the 13 errors showed that six were due to verification thresholds, which were set at the minimum of 0.48, a problem caused by the inconsistent sample signatures used in the reference design.

Discussion

The performance figures quoted in the previous section are conservative for several reasons. First, the results are completely unedited. The subjects themselves controlled the terminal with very little prompting from the experimenter. From time to time there were human errors, such as pausing in the middle of a signature, which stopped the recording, or advancing the paper tape at the wrong time, or writing too large for the space available and hitting the cover of the paper retainer. At least 30 such signatures were noted during the experiment, and it is likely that others passed into the data unnoticed. Therefore, at least 24 percent of the gross rejects were justifiable.

There was no training or practice. Most subjects required one to three trials to adjust to the strange pen and writing circumstances. This was quite noticeable in the overall time, where initial samples took much longer. The most direct effect of strictly taking the first five signatures for references was that the reference algorithm really had only three or four good signatures to deal with. On-line checking to assure five consistent samples is obviously needed. But in a sense, the human factors of taking a reference in one session are insurmountable. If the subject is not able to give a typical signature in the first session, there is not much to be done about it save finding a secure way to update the reference later, a so-called "post-enrollment" strategy of the type described in [6].

The subjects in these experiments were not greatly motivated nor very much distressed by a rejected valid signature. Also, although in real situations forgers receive a relatively high payoff, they perform under great stress. Our stress-free forgers appeared to be highly motivated by the gaming aspects of the experiment. Thus, we feel that both sides of the experiment are conservative; in an actual application, motivated signers will do better than our unmotivated subjects, and stressed forgers will do worse.

Regional correlation demonstrates an effective method for comparing signature dynamics. Previous techniques failed because they could not accommodate the normal distortions in the time axis (pauses, missing strokes, etc.). Regional correlation is computationally reasonable, not fine-tuned, and gives a familiar measure for the similarity of two signals. However, as our experience with real signatures and good forgeries increases, we expect to be able both to improve the performance and to simplify the computation by defining more powerful local measurements. It is also possible that more powerful techniques for measuring time distortions used in speech analysis may be applicable [16].

We found a considerable difference between the distributions of scores for valid signatures and forgeries for most of our subjects. One successful forgery was the result of too much variability in the design set, which caused the system to choose a low threshold. This is in accord with the experience of conventional document examiners. The forgeries that give trouble may be quite different from the reference, but the examiner may conclude that they are within the normal range of variation for this signer. Thus, improvements to the reference procedure are of equal value to increasing the sharpness of the decision rule.

The foremost technical problem remaining is to obtain consistency of performance for all subjects. It is conceivable that some individuals sign too inconsistently to use the system, but that is not the case with the data we have studied. The failures of our system are largely due to imperfections in the measurement design and instrumentation.

Finally, despite the many intrinsic human-factor problems we have mentioned, we believe our performance figures are sufficiently encouraging to indicate the feasibility of signature verification as a means of personal identification.

Acknowledgments

The authors thank their colleagues for their many important contributions. C. J. Evangelisti programmed the first implementation, wrote VPEN and the graphics support, and contributed his talents freely. N. Anthony designed and implemented the online system and was the prime mover in collecting the data. J. Morrissey helped formulate the initial idea. J. S. Birnbaum was instrumental in getting us started at a time when the idea was difficult to accept. The experimental dual-accelerometer pen was supplied by H. Baumeister and L. Larsen, IBM System Communications Division, Research Triangle Park, North Carolina.

References

1. W. Harrison, *Suspect Documents*, Praeger Publishers, New York, 1958, Chs. 9, 11.
2. O. Hilton, *Scientific Examination of Documents*, Callaghan and Company, Chicago, 1956, Chs. 8-13.
3. A. J. Mauceri, "Feasibility Study of Personnel Identification by Signature Verification," *Report No. SID 65 24 RADC TR 65 33*, Space and Information System Division, North American Aviation Co., Anaheim, CA, April, 1965.
4. R. F. Farag and Y. T. Chien, "Online Signature Verification," *Proceedings of the International Conference on Online Interactive Computing*, Brunel University, London, September, 1972, p. 403.
5. J. Sternberg, "Automated Signature Verification Using Handwriting Pressure," *1975 WESCON Technical Papers*, Paper No. 31/4, Los Angeles, September, 1975.
6. W. Haberman and A. Fejfar, "Automatic Identification of Personnel through Speaker and Signature Verification—System Description and Testing, "*Proceedings of the Carnahan Conference on Crime Countermeasures*, University of Kentucky, Lexington, May, 1976, pp. 23-30.
7. R. N. Nagel, "Computer Screening of Handwritten Signatures: A Proposal," *Report No. TR-220*, University of Maryland, College Park, MD, 1973.
8. T. P. Kabaservice and D. J. Muchow, "The Criminal Uses of False Identification—A Statement of a National Problem," *Proceedings of the Carnahan Conference on Crime Countermeasures*, University of Kentucky, Lexington, May, 1976, p. 1.
9. K. S. Lashley, "The Accuracy of Movements in the Absence of Excitation from the Moving Organ," *Amer. J. Physiol.* **20,** 169 (1917).
10. P. Bach-y-Rita, "Neurophysiology of Eye Movements," *The Control of Eye Movements*, Academic Press, Inc., New York, 1971.
11. J. van der Gon and J. Thuring, "The Guiding of Human Writing Movements," *Kybernetika* **1,** 145 (1965).
12. M. Eden and M. Halle, "The Characterization of Cursive Writing," *Information Theory*, C. Cherry, Ed., Butterworths Publishing Co., Washington, DC, 1961.
13. M. Yasuhara, "Steps Toward Handwriting Analysis," *Visible Language* **5,** 229 (1971).
14. J. Vredenbregt and W. G. Koster, "Analysis and Synthesis of Handwriting," *Philips Tech. Rev.* **32,** 73 (1971).
15. N. M. Herbst and J. F. Morrissey, U.S. Patent, 3,983,535, September, 1976.
16. F. Jelinek, "Continuous Speech Recognition by Statistical Methods," *Proc. IEEE* **64,** 532 (1976).

Received January 9, 1976

The authors are located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598.