Int. J. Human–Computer Studies (1995) **43**, 579–592

Optimizing digraph-latency based biometric typist verification systems: inter and intra typist differences in digraph latency distributions†

D. Mahar
School of Social Science, Queensland University of Technology, Beams Road, Carseldine, 4034, QLD, Australia

R. Napier and M. Wagner
The TRUST Project, RSISE, Australian National University, 0200, ACT, Australia

W. Laverty
Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon, ASAK, Canada, S7N 0W0

R. D. Henderson
Centre for Applied Psychology, University of Canberra, Canberra, 0200, ACT, Australia

M. Hiron
Department of Statistics, Australian National University, 0200, ACT, Australia

Umphress and Williams have shown that individual differences in digraph latency may provide a means of accurately verifying the identity of computer users. The present research refined this technique by exploring inter and intra subject differences in digraph latency distributions. Experiment 1 showed that there is marked heterogeneity in the latency with which individual subjects type different digraphs. Consequently, it was found that typist verification accuracy improved when a digraph-specific index of the distance between test and reference digraph latencies was employed. Experiment 1 also showed the utility of nonlinear modelling as a tool to establish optimum verification parameter settings. Experiment 2 showed that the use of a common low-pass temporal filter cutoff setting for all typists when screening digraphs is unwise. It was found that there is a significant interaction between subjects and filter settings such that verification accuracy may improve if subject-specific filter settings are used.

1. Introduction

The large quantity of both personally and organizationally sensitive data held on electronic media has led to the need for robust data security techniques. Cole (1978) has highlighted the importance of user verification as a means of ensuring data security. Current user verification paradigms include user knowledge (e.g. passwords), user artifacts (e.g. smart cards), and biometric systems which aim to identify the user via their physical or behavioural characteristics. This present research examined users' typing behaviour as a biometric method of verifying user identity, and thus enhancing computer system security. Although there is a range of potentially useful characteristics of users' typing behaviour, previous research

---

† This research has been carried out on behalf of the Harry Triguboff AM Research Syndicate, by The TRUST Project, Australian National University, Canberra, 0200, ACT, Australia.

1071-5819/95/100579 + 14$12.00/0

suggests that techniques based on inter-key latency measures are the most promising (Umphress & Williams, 1985; Leggett & Williams, 1988).

Current keystroke verification techniques can be broadly classified as either static or continuous. Static verification approaches analyse keystroke characteristic only at specific times, such as during the log-on process (e.g. Brown & Rogers, 1993; Joyce & Gupta, 1989; Newberry & Seberry, 1988, 1989). The verification error rates obtained with these static approaches have varied from 68% (Newberry & Seberry, 1988) to 14.1% (Brown & Rogers, 1993). Although the static approach provides more robust user verification than does the simple use of a password, it does not provide continuous security. For example, it cannot detect a substitution of the user after initial verification.

The continuous approach to user verification monitors the user's typing behaviour throughout the course of the interaction (e.g. Umphress & Williams, 1985; Leggett & Williams, 1988, Leggett, Williams, Usnick, & Longnecker, 1991). One such technique was developed by Umphress and Williams (1985), and later refined by Leggett and Williams (1988). A reference profile for each user is constructed consisting of the mean latency between the onset of successive keystrokes for each combination of keys (digraphs). The latency of digraphs typed during the test phase is then compared with the reference data for that digraph by measuring the proportion of test digraphs lying within a specified distance (expressed in terms of the standard deviation of all reference profile digraph latencies) of the relevant reference profile mean. Umphress and Williams (1985) report that verification accuracy is optimal when the maximum allowable separation between test and reference profile digraph latencies is set at 0.5 overall reference profile standard deviations and the proportion of test digraphs required to pass this test is set at 60%. Throughout both reference profile construction and test phases, all digraphs with latencies above a specified duration are excluded to control for cognitive breaks in typing. This type of user verification system has achieved false acceptance rates and false rejection rates of 5 and 5.5 percent respectively in a study using 36 subjects, a reference profile derived from 1075 keystrokes, and a test sample of 537 keystrokes (Leggett & Williams, 1988).

While the Umphress and Williams (1985) approach addresses a number of the problems of user verification, including the need for ongoing monitoring, their method has considerable potential for refinement. One limitation is the use of the standard deviation of all reference profile digraph latencies to assess the likelihood that a given test digraph latency comes from the authentic user's distribution of latencies for that specific digraph.

The pooled variance estimate only stands in a predictable relationship with the variance of a given digraph when there is homogeneity of variance across all reference digraph latencies. When this is not the case, the pooled variance estimate will frequently misrepresent the variability of the authentic user's individual digraph latencies. This will have the effect of reducing the false acceptance rate for digraphs with small variability at the expense of increasing the false rejection rate for digraphs with large variability. Conversely, the false rejection rate for digraphs with large variability will be reduced, while the false acceptance rate for digraphs with small variability will be increased. The overall effect of these variations on the total error of the procedure is dependent on how frequently each digraph is represented in the reference and test profiles.

One solution to this problem is to assess the difference between test and reference digraph latencies using the standard deviation of that particular digraph cell in the reference profile. By using this digraph-specific measure of variability the performance of the system becomes independent of both the degree of homogeneity of variance in latency between digraphs and the relative frequency of digraphs in the test and reference samples. While Umphress and Williams (1985) did explore the use such digraph-specific statistics, their typing samples did not include enough examples of each digraph to allow these statistics to be calculated reliably.

A further refinement can be made to the Umphress and Williams (1985) approach in the examination of optimum parameter values. The original technique employs two verification parameters; the maximum allowable separation between test and reference digraph latencies and the proportion of test digraphs passing this distance test. Neither the manner in which these parameters were optimized nor the effects that varying them have on the performance of the verification system were adequately addressed in the original studies (Umphress & Williams, 1985; Leggett & Williams, 1988). Hence, there is a need to clarify these two issues.

It is clear that the false acceptance rate will decrease as either the maximum allowable separation is decreased or the minimum proportion is increased. Conversely, the false rejection rate will decrease as either the maximum allowable separation is increased or as the minimum proportion is decreased. By manipulating these two parameters across a suitable range of values and then submitting the resulting false acceptance and false rejection rates to a process of nonlinear modelling, it is possible to predict the precise parameter settings where the combined false acceptance and false rejection rates are minimized. In addition, this process allows the specification of the region in this multidimensional space in which system performance falls within any specified range.

A third potential limitation of the existing digraph latency based technique is the use of a low-pass temporal filter set at the same value for all typists to exclude digraphs from analysis during both test and reference phases. The rationale of this approach is that digraphs with abnormally long latencies are not likely to be representative of the authorized users' typing. For example, these extreme latencies may represent cognitive breaks during which the typist is planning subsequent keystrokes. While this seems a reasonable proposition, it does not seem likely that one filter value is optimum for all typists. Although Leggett and Williams (1988) reported that a filter setting of 500 ms yielded optimum verification rates, Gentner (1983) found that the median inter-key latency of expert typists was 96 ms, while that of novice typists was 825 ms. Granted this, a 500 ms low-pass filter must exclude many keystrokes typical of a novice typist, while including many keystrokes which are not typical of an expert typist.

Clearly, this issue needs clarification by exploring the performance of the verification system as the filter setting is manipulated across a suitable range of values. In particular, the occurrence of a subjects-by-filter interaction would confirm that no one filter setting is optimum for all subjects.

The present research was therefore conducted with three main aims. First, it was designed to assess the degree to which the variances of the digraph latencies in each reference profile cell differed (i.e. the degree of heterogeneity of variance). If heterogeneity was present, then it was predicted that verification accuracy would be best when each test digraph was compared to the reference profile using a digraph-specific index of variability. The second aim of this study was to determine the relationships between the accuracy of verification and the two system parameters, and thus to specify the conditions under which optimum verification rates occur. Finally, this research set out to explore the extent to which verification performance varies as a function of low pass filter setting. The first two issues were explored in Experiment 1, while the issue of optimum filter setting was explored in Experiment 2.

2. Experiment 1: heterogeneity of digraph latency variance and nonlinear modelling of verification parameters

2.1. METHOD

2.1.1. Subjects

Sixty-seven subjects participated in the study. Fifty-eight were undergraduate psychology students while nine were TRUST project staff. Subjects were chosen regardless of typing skill, age or gender.

2.1.2. Design

A three-factor repeated measures design was employed. The first factor was the source of standard deviation used when test digraph latencies were compared with reference profile cell latencies. The standard deviation used was either the overall standard deviation of all reference profile digraph latencies or the standard deviation of the specific digraph under test.

The second factor manipulated the maximum allowable separation between test and reference digraph latencies; that is how many standard deviation away from the reference profile mean a given test digraph was allowed to be before being judged as atypic of the authorized user. This factor was varied between 0.05 and 1.0 standard deviations in steps of 0.05 and from 1.0 to 5.0 standard deviations in steps of 0.5.

The third factor was the minimum proportion of test digraphs which were required to pass the verification test before the complete test string was judged to be typical of the authorized user's typing. This factor was varied between 5 and 95% in 5% steps. The dependent measures were the false acceptance rate and false rejection rate observed for each subject in each condition.

2.1.3. Procedure

All subjects completed a transcription typing task in which they were presented with textual stimuli on a computer monitor then typed the stimuli into a dialogue box on the bottom half of the screen. The subjects initially undertook a practice session in which the stimuli consisted of four words, four sentences and one name. The subjects were allowed as much practice as they desired before commencing the experimental trials. This ensured that all subjects were familiar with the experimental paradigm prior to data gathering.

The stimuli in the experimental session were 30 high frequency words, including common computer control commands, which were each presented twice; once in isolation, and once within a sentence. Ten of the words were two characters long, 10 were four characters long, and 10 were six characters long. There were also six presentations of a key sequence familiar to the subject. In some cases this included the subjects name, but in other cases it consisted of other words of equivalent length. These 72 stimuli were randomly ordered throughout the experimental session. If the subjects did not copy a given stimulus accurately then that trial was repeated at the end of the session until all the stimuli had been typed correctly.

As the subjects typed the text, the key-down time for each keystroke was recorded at a resolution of 1 ms. A reference profile was then constructed for each subject from a random sample of 500 digraphs drawn from their text. This sample included only combinations of lower-case alphabetic keys and the space bar. Other combinations were judged to represent either typing errors (e.g. those including editing keys) or the simultaneous use of two keys (e.g. upper case characters). To ensure the robustness of the descriptive statistics calculated for each reference profile cell, only those digraphs for which at least five observations were available were included in the reference profile. Similarly, only those digraphs with a latency less than 500 ms were included to control for cognitive breaks.

On average, this screening process yielded reference profiles containing 21.36 useable cells, with 85% of the profiles including at least 15 useable cells. The 10 most frequent digraphs were; r-e, e-space, t-o, t-h, t-space, i-n, s-space, h-e, space-t, and space-s. The average number of observations available in each useable reference profile cell was 7.62 digraphs. Once the digraphs were assigned to the reference profile, the mean latency and standard deviation for each useable cell were calculated, along with the standard deviation of all reference profile digraph latencies.

Test data were then compared with the established reference profile to derive false rejection and false acceptance rates. The false rejection rate for each subject was calculated using 66 random samples of 100 digraphs drawn from the pool of keystrokes not used to construct the reference profile. Each test digraph latency was compared with the corresponding reference profile cell mean digraph latency as the maximum allowable separation was varied from 0.05 to 5.0 standard deviation. This process was undertaken using both the digraph-specific and overall reference profile standard deviations as the distance measure. In each case, the minimum proportion of test digraphs passing this distance test before verification was varied from 5% to 95%. The false rejection rate for each subject in each condition was defined as the percentage of the subjects' 66 test samples not accepted as legitimate.

False acceptance rates were calculated by testing a random set of 100 digraphs drawn from each subject's pool of keystrokes against the reference profile of each other subject. The procedure followed was the same as that used to generate the false rejection data, except that the false acceptance rate for each subject was defined as the proportion of "impostor" samples which this test failed to exclude as invalid. This process was repeated at each level of the three experimental factors.

2.2. RESULTS AND DISCUSSION

The homogeneity of digraph cell variances was examined for each subject using Hartley's $F_{\max}$ test (Kirk, 1982) to compare the variance of the most and least variant cell in the subject's reference profile. This analysis revealed that in all but two of the 67 cases there was substantial heterogeneity of digraph latency variance as indicated by $F_{max} > 9$. This result strongly supports the argument that the use of a pooled variance estimate is not statistically appropriate.

A subset of the false acceptance and false rejection data was submitted to repeated measures MANOVA in order to confirm that the manipulation of all three factors did alter the accuracy of verification. For this analysis the maximum allowable separation factor was set at either 0.25, 0.5, 1.0, 2.0 or 4.0 standard deviations, while the proportion of digraphs required to pass the verification test was set at 20, 40, 60, or 80%. This analysis revealed significant main effects due to the source of standard deviation, $F(2, 65) = 401.99$, $p < 0.001$, maximum allowable separation, $F(8, 59) = 72.77$, $p < 0.001$, and minimum proportion, $F(6, 61) = 25.41$, $p < 0.001$, factors. However, significant interactions were also found between the source of standard deviation and maximum allowable separation, $F(8, 59) = 5775.96$, $p < 0.001$, source of standard deviation and minimum proportion, $F(6, 61) = 1010.25$, $p < 0.001$, and maximum allowable separation and minimum proportion, $F(24, 43) = 40.10$, $p < 0.001$, factors. Finally, the three way interaction between these factors was also found to be significant, $F(24, 43) = 740.52$, $p < 0.001$. The significant interactions between these factors indicate that verification rates are not a simple function of the three factors.

The nature of these interactions between the factors was explored via a process of nonlinear modelling. Two separate models of the relationship between mean verification error rates and the two verification parameters were generated. The first model related the false acceptance and false rejection rates (expressed as FA + FR†) in the digraph-specific standard deviation condition to the maximum allowable separation and minimum proportion parameters. The second model repeated this process using the data from the overall reference profile standard deviation condition. Each model accounted for greater than 99.7% of the variance in the respective error rate data, with a random distribution of residuals suggesting that only error variance remained.

Figures 1 and 2 present surface plots depicting the shape of these functions for each condition. While both functions show poor levels of verification accuracy over most of the range of verification parameter settings, a very rapid rise in verification accuracy is apparent as these parameters approach their optimum settings. In both cases, the region of near optimal performance is not constrained to a limited set of verification parameter settings. That is, for any given value of one verification parameter, there is a corresponding value of the other parameter which together yield near optimum performance.

The digraph-specific method yielded a minimum predicted verification error rate of 35%‡ at a maximum allowable separation value of 1.63 standard deviations and a minimum proportion passing setting of 73%. These predicted values conformed well with the minimum observed verification error rate of 35% obtained at a maximum separation value of 1.5 standard deviations and minimum proportion setting of 70%. The overall reference profile condiiton yielded a higher predicted verification error rate of 43% when the maximum allowable separation parameter was set at 0.86

---

† This metric weights FA's and FR's as being equally important. Depending upon the requirements of the particular setting, real world applications might choose to weight these two types of errors differently.

‡ In this section, all error rates are expressed as FA + FR.

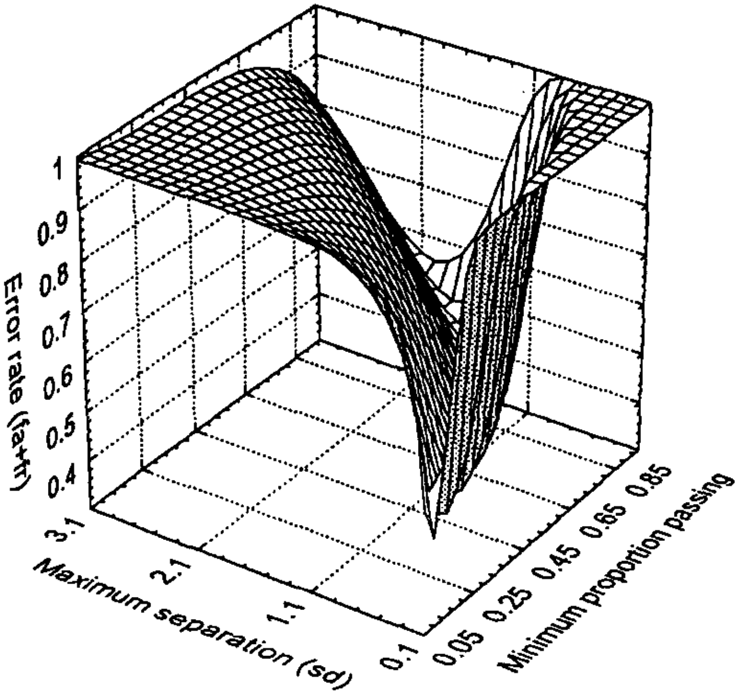

FIGURE 1. Surface plot depicting the relationship between verification error rate (FA + FR) and both the maximum allowable separation between test and reference digraph latencies and the minimum proportion of test digraphs passing verification using the overall reference profile index of digraph latency variability.

standard deviations and the minimum proportion passing parameter was set at 65%. These values were also similar to those observed in the raw data, where a minimum verification error rate of 38% was obtained at a maximum separation setting of 0.85 standard deviations and a minimum proportion passing setting of 65%.

The superior performance of the digraph-specific condition was not constrained to a narrow range of verification parameter settings. Figure 3 shows the wide range of settings over which this technique yielded superior performance to the optimum performance observed at any point using the overall reference profile based distance measure.

The primary finding of this study was that there is marked heterogeneity of variance in the latency with which individual subjects type different digraphs. This result highlights that Umphress and Williams' (1985) use of the pooled estimate of digraph variability is inappropriate. The results also demonstrated that this problem of heterogeneity of variance exerts a measurable effect on the accuray of user identity verification. In particular, the total verification error rate was found to fall from 38% under the Umphress and Williams (1985) approach to 35% when a digraph specific estimate of variability was used. This superiority was found to hold over a wide range of system settings, thus further confirming the need to employ a digraph-specific variance estimate.

Second, the present study clarified the effects of manipulating the maximum allowable separation and minimum proportion of test digraphs passing verification

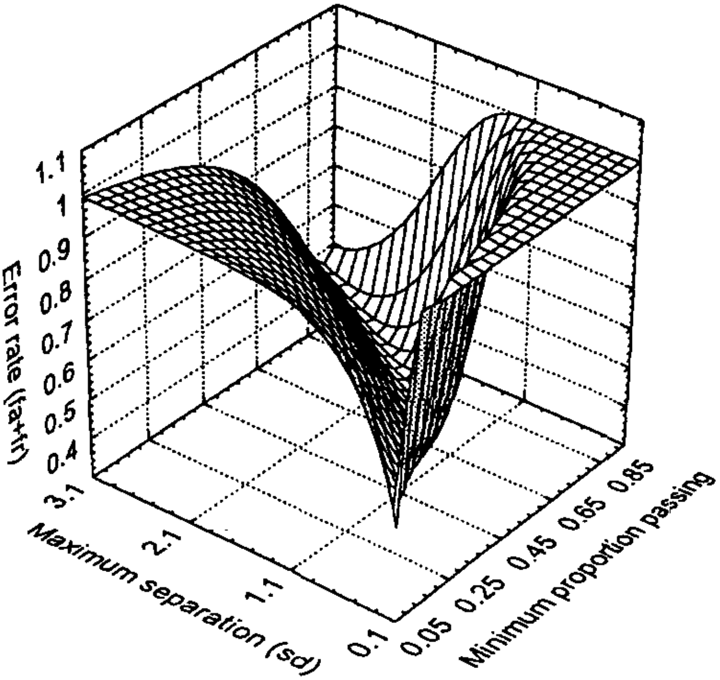

FIGURE 2. Surface plot depicting the relationship between verification error rate (FA + FR) and both the maximum allowable separation between test and reference digraph latencies and the minimum proportion of test digraphs passing verification using the digraph-specific index of digraph latency variability.

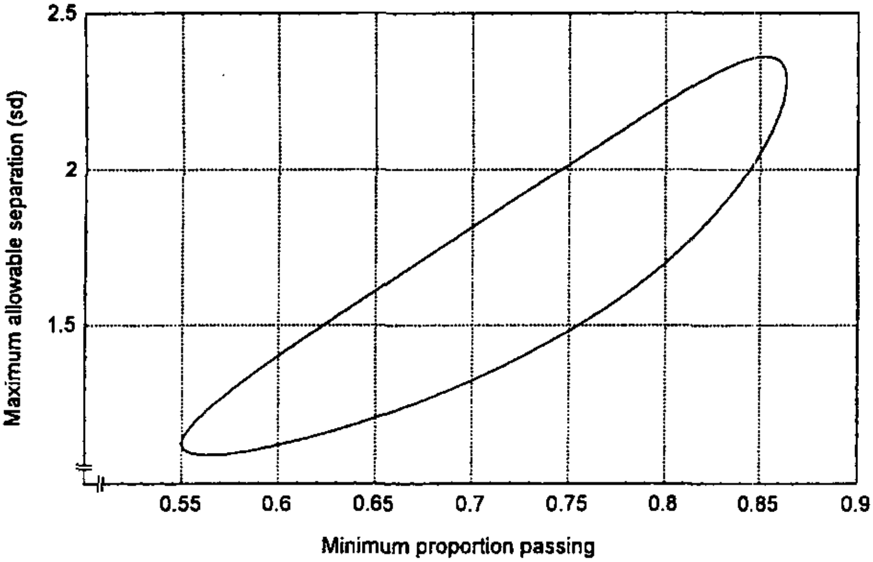

FIGURE 3. Contour plot showing the range of verification parameter settings over which verification error rates (FA + FR) using the digraph-specific index of digraph latency variability were superior to the optimum verification error rate achieved using the overall reference profile index of digraph latency variability.

parameters on the resulting total verification error rate. Analysis via MANOVA confirmed that both these factors caused significant change in the total verification error rate. However, the occurrence of significant interactions between these parameters shows that the relationship between these parameters and verification accuracy is complex.

Nonlinear modelling of the relationship between these parameters and verification accuracy yielded models which accounted for more than 99.7% of the variance in verification rates. While the analysis showed that the performance of the verification system was relatively stable at near optimum levels across a wide range of settings, the models also identified the most efficient settings of the maximum separation and minimum proportion passing parameters. In the case of the overall standard deviation condition these optimum settings are similar to those achieved in previous studies (Umphress & Williams, 1985; Leggett & Williams, 1988), suggesting that, while total verification erorr rates may differ across studies, the optimum verification parameter settings are stable.

In spite of these promising results, it must be noted that the total verification error rates observed in this study are substantially higher than those reported elsewhere (Umphress & Williams, 1985; Leggett & Williams, 1988). There are, however, a number of factors which may account for these differences. First, the present study used substantially smaller test samples than were used in those earlier studies, thus reducing its relative sensitivity. Second, a much more stringent technique was used here to determine verification error rates. In particular, the present false rejection rates were calculated across 66 test samples rather than from a single test sample as used by Umphress and Williams (1985). Finally, the present study tested a substantially larger sample of typists than previous studies, hence increasing the likelihood that the sample would include subjects with similar typing patterns. None of these observations challenges the key finding that digraph latencies are markedly heterogeneous, and consequently that a digraph-specific measure of variability is required.

3. Experiment 2: verification accuracy as a function of low-pass filter setting

This experiment was conducted to assess whether Leggett and Williams' (1988) claim that a 500 ms low-pass temporal filter provides optimum verification rates is correct. In lines with Genter's (1983) data, it was predicted that no one filter setting would prove optimum for all subjects. Although Experiment 1 provided extensive information regarding the optimum setting of the maximum allowable separation and minimum proportion verification parameters, these factors were included in the present study in order to detect any interaction between them and the filter setting.† Finally, granted the results of Experiment 1, a digraph-specific estimate of digraph latency variability was used in this study.

3.1. METHOD

3.1.1. Subjects

Data from the same 67 subjects tested in Experiment 1 were used here.

---

† Extra factors were also needed to allow the construction of a suitable error term to test the subjects-by-filter interaction.

3.1.2. Design

This experiment employed a three-factor repeated measures design. The first factor was the low-pass filter setting used. This was varied from 300 to 800 ms in 100 ms steps. The second factor was the maximum allowable separation between test and reference digraphs. This factor was varied between 0.5, 1.0, 2.0, and 4.0 standard deviations. The final factor was the proportion of test digraphs required to pass the maximum allowable separation test before verification was assumed. This factor was varied from 20 to 80% in steps of 20%. The dependent variables were the false acceptance and false rejection rates for each subject in each condition.

3.1.3. Procedure

With the exception of the particular factors manipulated, the procedure followed in this experiment was identical to that used in the digraph-specific condition in Experiment 1.

3.2. RESULTS AND DISCUSSION

The false acceptance and false rejection rates for each subject at each level of the filter, maximum separation, and minimum proportion factors were averaged to provide an overall index of verification accuracy in each condition. All subsequent analyses were based on these average error rates. Figure 4 depicts cell means as a function of these three factors. At all filter settings, the average optimum verification rate was just under 20%. As per Leggett and Williams' (1988) findings, the best

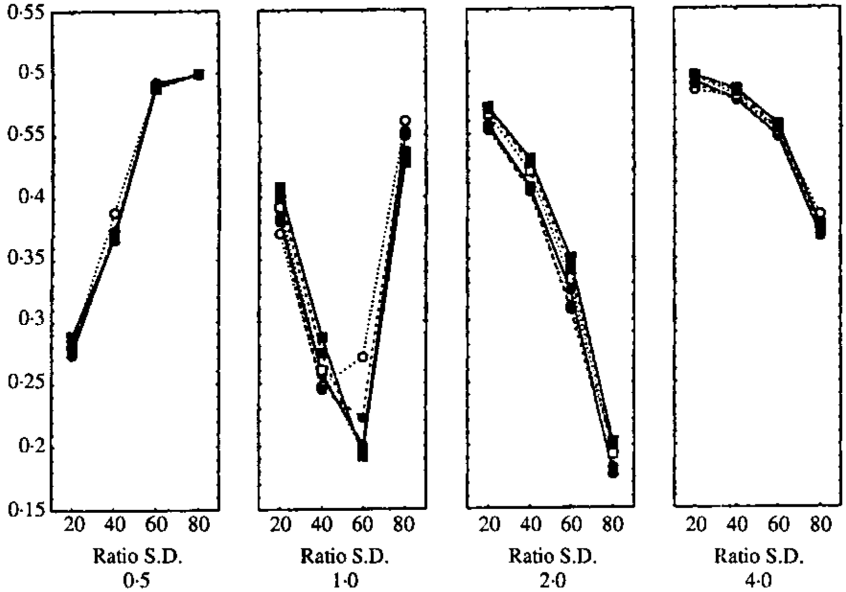

FIGURE 4. Interaction plot showing the relationship between verification error rate (FA + FR), the low-pass temporal filter setting, the maximum allowable separation between test and reference digraph latencies, and the minimum proportion of test digraphs passing verification. Filter key:- · · ·○· · ·: 300 ms; – – ● – – –: 400 ms; —●—: 500 ms; · · · □ · · ·: 600 ms; · · ·■· · ·: 700 ms; —■—: 800 ms.

verification rate (17.6%) was obtained with a filter setting of 500 ms. As the filter setting was varied away from 500 ms, optimum verification rates tended to fall off progressively; although only by 1 to 2%. Granted these small differences, it is clear that the cutoff setting for the low-pass temporal filter did not have a strong effect on the accuracy of verification.

The settings of the maximum allowable separation and minimum proportion parameters exerted a stronger effect on verification rates, although, as in Experiment 1, these two factors appeared to interact strongly. In particular, as the maximum allowable separation between test and reference digraph latencies was increased, the minimum proportion setting at which optimum verification performance occurred progressively increased. For example, when a maximum allowable separation setting of 0.5 standard deviations was used, optimum verification occurred at a minimum proportion setting of 20%. However, at a maximum allowable separation setting of 4.0 standard deviations, optimum verification rates occurred at a minimum proportion setting of 80%.

Finally, Figure 4 indicates that a three way interaction occurred between these factors. In particular, optimum performance was observed with a maximum allowable separation setting of two standard deviations and a minimum proportion passing setting of 80% when the filter cutoff was set at 500 ms or less. However, at filter cutoffs greater than 500 ms, optimum performance was observed with settings of one standard deviation and 60% respectively. In both cases, these settings are similar to the optimum maximum allowable separation and minimum proportion settings determined in Experiment 1 (1.67 S.D. and 73% respectively).

A $6 \times 4 \times 4$ repeated measures ANOVA was conducted to confirm these observations. As the sphericity assumption was breached in all cases, the treatment and error degrees of freedom were adjusted to control the Type I error rate (see Myers & Well, 1991; p. 248). While no significant main effect of filter setting was observed, $F(5, 330) = 1.64$, $p > 0.05$, significant effects due to both the maximum allowable separation, $F(1, 66) = 163.16$, $p < 0.05$, and the minimum proportion, $F(1, 66) = 36.56$, $p < 0.05$, factors were found. In addition, the two-way interactions between maximum allowable separation and minimum proportion, $F(1, 66) = 222.73$, $p < 0.05$, that between maximum allowable separation and filter cutoff, $F(3, 241) = 3.17$, $p < 0.05$, and that between minimum proportion and filter cutoff, $F(1, 66) = 9.10$, $p < 0.05$, were all found to be significant. Finally, the three-way interaction between these factors was also found to be significant, $F(1, 66) = 5.08$, $p < 0.05$.

While on average the best performance was achieved using Leggett and Williams' (1988) 500 ms low-pass temporal filter, this does not mean that this setting is best for all subjects. As was argued earlier, Gentner's (1983) keystroke latency data suggests that a set filter might not be appropriate for all levels of typing skill. In order to assess this possibility descriptively, the distributions of the latencies of the ten most frequent digraphs in the reference profiles were extracted for each subject. Figure 5 presents box plots for two selected subjects to highlight the outcome of this analysis. In the case of Subject 23, it can be seen that the latencies of all but the most extreme digraphs lie below 500 ms. However, in the case of Subject 24, even the most extreme digraphs have latencies less than 500 ms. Equivalent data from the other subjects revealed a variety of shapes of digraph latency distribution, some even

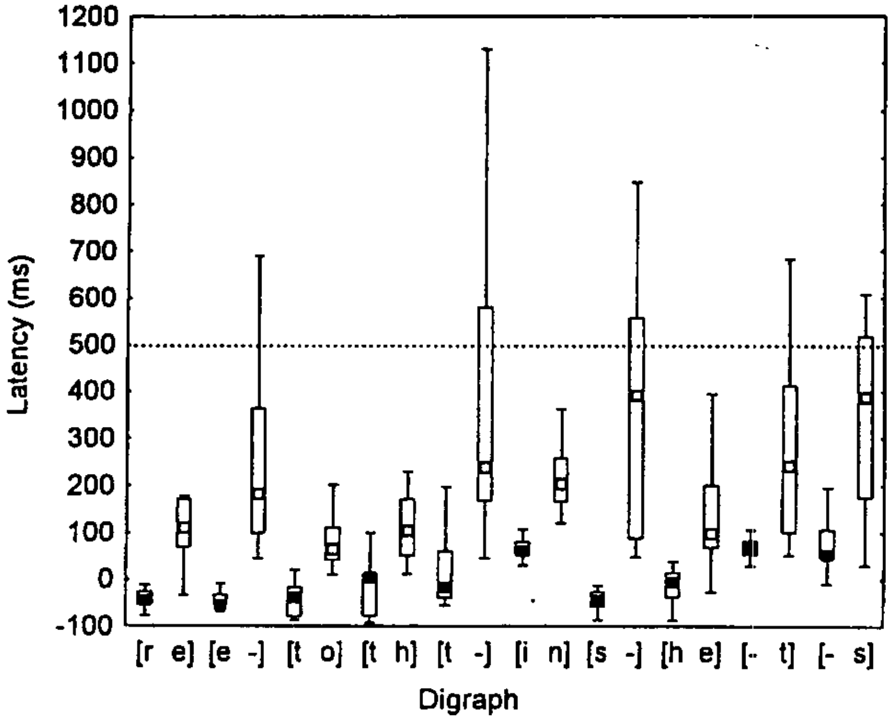

FIGURE 5. Box plot showing the distribution of digraph latencies for the 10 most frequently occurring digraphs in the reference pofiles of Subjects 23 and 24. The symbol "–" denotes the "space" character.

more extreme than these two examples. Clearly, these results suggest that optimum verification rates for each subject will not be achieved using a common filter setting.

An inferential test of the extent to which the inter-subject variations in digraph latency influenced verification accuracy was obtained by assessing the significance of the subjects-by-filter interaction. This analysis, again using a conservative F-test to protect against breaches of sphericity, revealed a significant interaction between subjects and filter setting, $F(1, 66) = 5.79$, $p < 0.05$.†

4. Conclusions

In summary, these two experiments have highlighted the utility of the Umphress and Williams' (1985) digraph latency approach to dynamic typist verification. In addition, they have revealed two areas in which this approach may be improved. Experiment 1 showed that there is marked heterogeneity in the variability with which typists produce each digraph, and hence that the use of a pooled estimate of digraph latency variability is inappropriate. Indeed, it was found that the use of a digraph-specific index of latency variability led to measurable improvements in verification accuracy across a range of other parameter settings. Finally, Experiment

---

† In this case, the interaction treatment mean square is given by the error term used to test the main effect of the filter factor, while the appropriate error mean square is given by the error term used to assess the three-way interaction between filter, maximum separation, and minimum proportion.

1 showed the usefulness of applying nonlinear modelling when attempting to determine the optimum settings of verification parameters like the maximum allowable separation between test and reference digraphs and the minimum proportion of test digraphs passing verification.

Experiment 2 confirmed that a 500 ms low-pass temporal filter does, on average, yield optimum verification rates, but also showed that better results may be obtained using subject-specific filter settings. This approach could be taken even further, by applying different filter settings for each digraph in each subject's reference profile. Indeed, although Figure 5 shows considerable overlap between digraph latency distributions within each subject, there are discernible differences in the distributions for each digraph. Clearly, subsequent research should explore these possibilities by comparing verification accuracy using set filter values, subject-specific filter values, and subject-by-digraph-specific filters.

Beyond the results of the present experiments, there are other ways in which the Umphress and Williams (1985) approach might be improved. For example, the choice of key-down to key-down time as the index of typing behaviour may not be optimal. This interval involves two potentially orthogonal components; the duration that the first key is depressed, and the period between the release of that key and the depression of the second key. It may be that pooling these two components masks useful individual difference information. Indeed, some recent research has explored the use of a multivariate verification strategy (Brown & Rogers, 1993) in the context of a neural network based verification strategy.

Of course, there are practical problems in the application of either continuous or static typist verification systems. Firstly, it is possible for a user to do much damage with only a few keystrokes (for example "del*.*"). Thus further refinement, perhaps along the lines indicated above, is required to ensure that the best possible verification rates are achieved with the briefest possible test strings. Once this limiting test string length is known, it may be necessary to alter the length of all "high security" commands to at least this length. An alternative to this, which is inherent in the concept of continuous verification, is to pool samples of the user's typing across a moving temporal window covering more that the current command, but with greater weighting being given to the most recent keystroke events. Second, there has been a rapid increase in the use of graphical user interfaces in which little typing is required. Indeed, there has been research undertaken to develop verification techniques based on the modes of input frequently employed in these new interfaces (i.e. pointing, gesturing, writing, and speech). These developments do not negate the utility of typing verification systems. Rather they highlight the need for adequate verification paradigms to deal with whichever mode of input the user elects to use.

References

Brown, M. & Rogers, S. J. (1993). User identification via keystroke characteristics of typed names using neural networks. *International Journal of Man–Machine Studies*, **39**, 999–1014.

Cole, D. (1978). *Design alternatives for computer security.* NBS Special Publication. pp. 500–521. Washington, DC: US Government Printing Office.

GENTNER, D. R. (1983). Keystroke timing in transcription typing. In W. I. COOPER, Ed. *Cognitive Aspects of Skilled Typewriting.* pp. 95–120. New York, NY: Springer-Verlag.

JOYCE, R. & GUPTA, G. (1989). *User authorization based on keystroke latencies.* Technical Report. Department of Computer Science, James Cook University, Queensland Australia.

KIRK, R. E. (1982). *Experimental design: procedures for the behavioural sciences.* Pacific Grove, CA: Brooks Cole.

LEGGETT, J. & WILLIAMS, G. (1988). Verifying identity via keystroke characteristics. *International Journal of Man–Machine Studies,* **28,** 67–76.

LEGGETT, J., WILLIAMS, G., USNICK, M. & LONGNECKER, M. (1991). Dynamic identity verification via keystroke characteristics. *International Journal of Man–Machine Studies,* **35,** 859–870.

MYERS, J. L. & WELL, A. D. (1991). *Research Design & Statistical Analysis.* New York, NY: Harper Collins.

NEWBERRY, M. & SEBERRY, J. (1988). *User unique identification.* Technical Report CS88/12, University College, University of New South Wales, Australia.

NEWBERRY, M. & SEBERRY, J. (1989). Experience of using a type signature password system for user authentication in a heavily used computing environment. In W. J. CAELLI, Ed. *Computer Security in the age of Information.* pp. 303–307. Amsterdam: Elsevier Science Publishers.

UMPHRESS, D. & WILLIAMS, G. (1985). Identity verification through keyboard characteristics. *International Journal of Man–Machine Studies,* **23,** 263–273.