Interacting with Electronic Work Surfaces: Studies in Note Taking and Writing Behavior

Principal investigator: Albert N. Badre

Georgia Institute of Technology

Final Report

December 31, 1990

Effects of Long Term Usage

Albert N. Badre
Paulo Santos
Georgia Tech

Handwriting recognition systems are to be used by both novice and more experienced users. Every new user of such a system will initially be a novice, with no previous experience of using the system, and will gradually become more and more proficient in using a system by using the acquired expertise. Several experiments have been executed to study the influence of several effects on handwriting recognition, user satisfaction, and overall system performance. However, due to the short duration of all those experiments, all the data collected thus far pertains exclusively to novice users. In order to uncover effects of long-term usage, another experiment was designed and performed. This report presents the results of such an experiment.

This experiment is the sixth and last in a series of experiments investigating the effect of several variables on the performance of a handwriting recognition system.

Purpose

The primary goal of this experiment was to investigate the effects of long term usage of a handwriting recognition system in user behavior and satisfaction, and in system performance. The hypothesis we formulated was that users would change their behavior in order to increase the system recognition accuracy as they become more experienced at using the system. Secondary purposes of the experiment were to identify the ways in which people change the ways they form characters, and to identify possible uses of the timing information to improve recognition accuracy.

Description of the experiment

The experiment consisted of capturing handwriting samples from subjects, writing on a transparent digitizing tablet. Fourteen subject, all students at Georgia Tech, were asked to visit the usability laboratory at Georgia Tech a total of fourteen times over a period of two months, with no more than a session per day. Laboratory sessions were scheduled by appointment, and most subjects maintained a regular schedule of two visits per week. Two feedback conditions were being studied: immediate continuous feedback, where the subjects received feedback on recognition concurrently with their writing task; and postponed feedback, where subjects received feedback at the end of each page of text.

The pool of subjects consisted of 14 Georgia Tech students, ten males and four females, with ages ranging from 21 to 45 years and averaging 26 years. Thirteen of the subjects were right-handed, and one (male, 23) was left-handed. Subjects were paid for their participation in the experiment, and their pay was prorated to the number of characters recognized by the system in each session. They would also be given an extra bonus for completing all fourteen sessions - which they all did - that was also prorated to their average performance over all fourteen sessions.

Before the first session their were briefed on the purpose and nature of the experiment, and on their duties as subjects. They read and signed a release form, informing them of their rights (including quitting the experiment at any time) and allowing us to collect the required data. They were also instructed on how to operate the system, and were shown a sheet with guidelines on how to form characters. They were allowed to refer to the guideline sheet at any time during the sessions. At the end of the first session they were asked to fill out a questionnaire with demographic information. Furthermore, at the end of the first, seventh and fourteenth sessions, they were asked to fill out a questionnaire with their opinions on the system.

Each session consisted of writing a text, which was projected onto the wall, into a computer, using a transparent digitizer which rested above an LCD. Each subject was instructed to write in all block, uppercase letters (the only type of print recognizable by our algorithm) and to leave a small space between letters. They wrote in the bottom half of the digitizer, which presented a lined display format, with lines spaced about 7mm apart. The text consisted of alphabetic characters, digits and punctuation, and the text length ranged from 419 to 638 characters (mean 563, standard deviation 67). In each session, the text was presented as a sequenced of four transparencies projected onto the wall in front of the subjects. The overall distribution of characters in the paragraphs is shown in table 1.

A	6.98%	B	1.32%	C	2.87%	D	3.19%	E	12.04%	F	2.52%
G	1.72%	H	4.02%	I	7.22%	J	0.25%	K	0.61%	L	3.60%
M	2.73%	N	6.05%	O	7.45%	P	1.89%	Q	0.25%	R	6.16%
S	6.25%	T	7.72%	U	3.04%	V	1.13%	W	1.68%	X	0.25%
Y	2.10%	Z	0.23%	0	0.65%	1	0.65%	2	0.30%	3	0.26%
4	0.26%	5	0.31%	6	0.21%	7	0.34%	8	0.18%	9	0.47%
		Punctuation			3.1%						

Table 1 - Distribution of characters over all texts

Two independent variables were present in this experiment: session number (1 through 14) and feedback (postponed or continuous). This resulted in a 14x2 2-factor experiment design, with a total of 28 conditions. In the postponed feedback condition, subjects saw what the system recognized at the end of each screen full of text, which generally corresponded to one page of text projected onto the wall. In the continuous feedback condition, the text being recognized by the system was being displayed simultaneously (or with a maximum delay of one second) with the user's writing. In this case, the users wrote in the bottom half of the screen and the recognized text appeared in the top half of the screen.

Equipment and recognition overview

Subjects wrote on a Seiko D-Scan digitizer/display device. The display was passive matrix VGA with approximate dimensions of 289mm wide X 157mm high. It was covered by a transparent digitizer and a glass plate, and connected to an NCR 386 PC. A stylus that was tethered to the digitizer was used as the writing device.

Stroke data collected from the digitizer device were fed to a rule-based character recognition system based on feature detection. Strokes were made up of a series of points sampled from the digitizer at a constant frequency of 143 Hz during pen down states. The recognition software attempted to both segment (identify which strokes should be grouped to form a single character) and recognize (identify which character the strokes represent) the stroke data. For recognition to occur, a set of strokes must be segmented correctly and pass all the rules for one and only one character in the system's rulebase. Otherwise, an "unrecognized" response is returned for a set of strokes. Once strokes are segmented to contruct a character (either correctly, incorrectly, or yielding an unrecognizable situation), they are considered used and are not included in future attempts to segment and recognize characters. Therefore, an incorrect segmentation of a character may lead to a series of incorrect recognitions due to incorrect segmenting.

Factors affecting timing performance

The effects of the two factors involved in this experiment on the timing performance were analyzed. Two dependent variables were considered: the overall time to complete a session, and the total pen-down time to complete a session. The overall time to complete a session is computed as the sum of the times taken to complete a pages, for every page in a session. This method ignores the time spent between pages, but it was nevertheless chosen because the time between pages carries a large variability due to factors that are not object of the study of this experiment. As a consequence, it is not possible to determine the time spent by those subjects receiving postponed feedback, which occurs between pages.

The effects of practice on timing performance

Practice at using the system is a direct result of repeated usage. Therefore, higher practice levels can be equated with larger session numbers, and vice versa. The formulated hypothesis is that the time to write decreases with practice. Stated in different terms, the hypothesis states that there is a negative correlation between the session number (practice) and the tim required to write a fixed number of characters. Figure 1 shows the average time required to write 1000 characters, and a trend for the negative correlation is evident. Statistical significance is found in the completion time decrease between sessions that are 10 session numbers apart, such as between 1 and 11 or between 2 and 12.

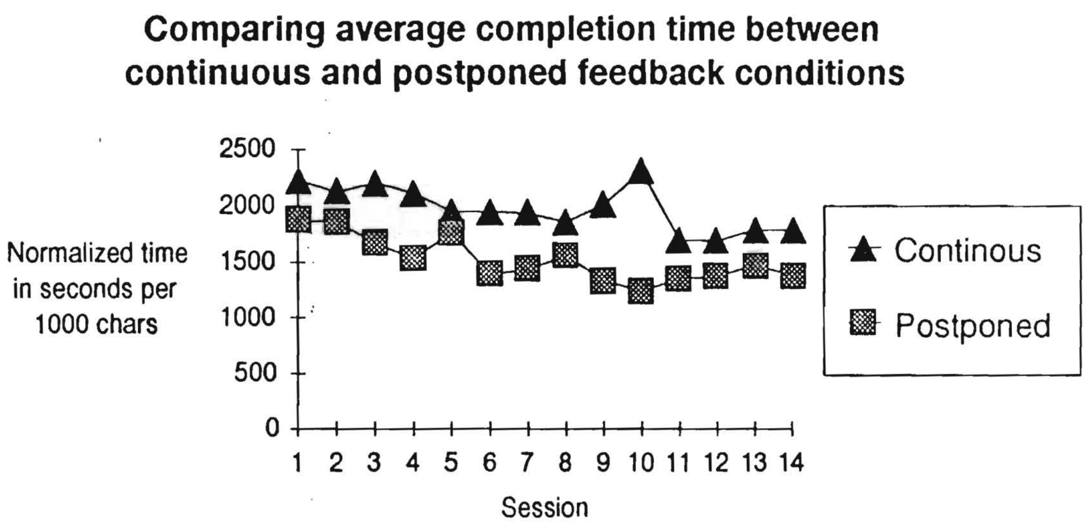

Figure 1

A quick glance at figure 1 suggests that users perform faster when using postponed feedback than when using continuous feedback. Let us clarify that this is not necessarily true, and how the above data can not be interpreted in that way. When users are performing their task with the system, they are either reading text to be written, writing that text, or checking the correctness of their writing. However, in the postponed feedback condition they only check the correctness after the end of each page of input. The way in which the experimental system calculates the total time to complete the task is as the sum of the times to complete each page. The time to complete each page is defined as the time that elapses between the instant the user first puts the stylus down on that page, and the instant the stylus is lifted up after the last character on the page. Therefore, the time required to check the feedback in the postponed feedback condition is not being considered, and this may account for a significant time that is not quantified in this experiment.

A comparison of the writing times can be made by comparing the total time that the pen is down in each session. This should not be considered as the time to complete a writing task, but it illustrates that there does not seem to be any significant difference in the writing times between postponed and continuous feedback conditions. Figure 2 shows the comparison of total pen-down time for each session, under the two feedback conditions.

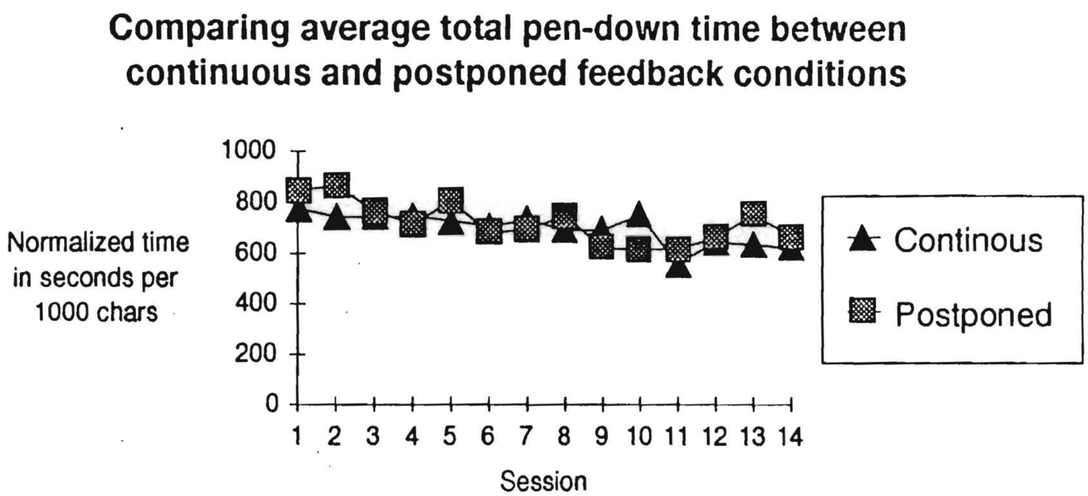

Figure 2

By analyzing the data displayed in figure 2, we can not claim that the type of feedback (continuous or postponed) has an effect on total pen-down time: the time required to actually write. We can still claim, however, that there are is a negative correlation between practice and total pen-down time. Statistical significance is again found in the completion time decrease between sessions that are 10 session numbers apart, such as between 1 and 11 or between 2 and 12.

Timing analysis of the formation of characters

An extensive analysis was performed on the temporal information associated with character formation. Each recognized character consisted of a sequence of one to four strokes with the stylus on the digitizer. Of all the 100,000+ characters captured during the experiment, 46.2% were formed with a single stroke, 29.4% had two strokes, 20.4% three strokes, and 4.0% four strokes. The average times to form one-stroke characters was 555 milliseconds; for two-stroke characters, 916ms; for three-stroke characters, 1117ms; and for four-stroke characters, 1551ms.

Pen-up times within and between characters

One of the most significant problems in recognition is the proper segmentation of characters. The CIC algorithm used in this experiment performs the segmentation of characters based exclusively on spatial characteristics of the characters. In previous experiments we had empirically verified that one of the most frequent causes of error was poor segmentation. Therefore, we investigated whether stroke timing information could be of any usefulness for the segmentation problem.

The problem with segmentation occurs due to the existence of multiple stroke characters. For correct segmentation, an algorithm has to group correctly the set of strokes that forma character, and to separate these strokes from those that belong to the preceding and following characters. Our hypothesis is that shorter pen-up times occur within the same character than between separate characters. To confirm or deny this hypothesis, we analyzed all the pen-up time intervals that occurred between strokes of the same character, and those that occurred between strokes belonging to two characters of the same word. Figure 3 shows the distribution of such times, for intervals shorter than 1 second and with a 55 millisecond granularity.

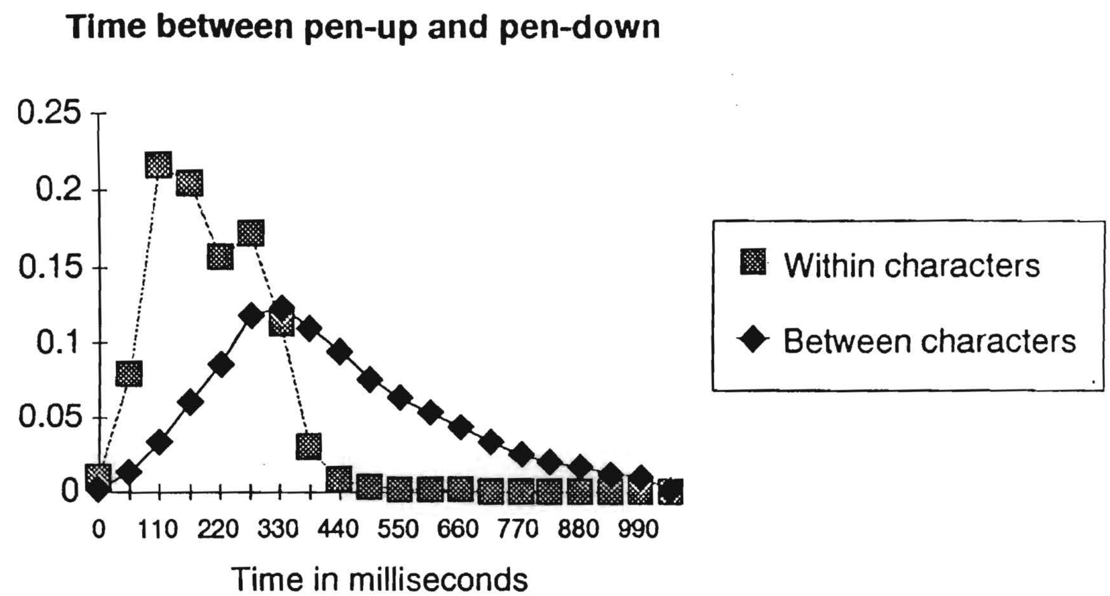

Figure 3

Figure 3 shows, as expected, that strokes that occur within the same character are closer apart in time than strokes that belong to different characters. The significance of this result is increased when we ask the question from a different perspective: Given an pen-up time *t*, what isd the probability that the two strokes separated by that pen-up belong to the same character? Figure 4 shows the answer that is derived from the expeerimental data collected.

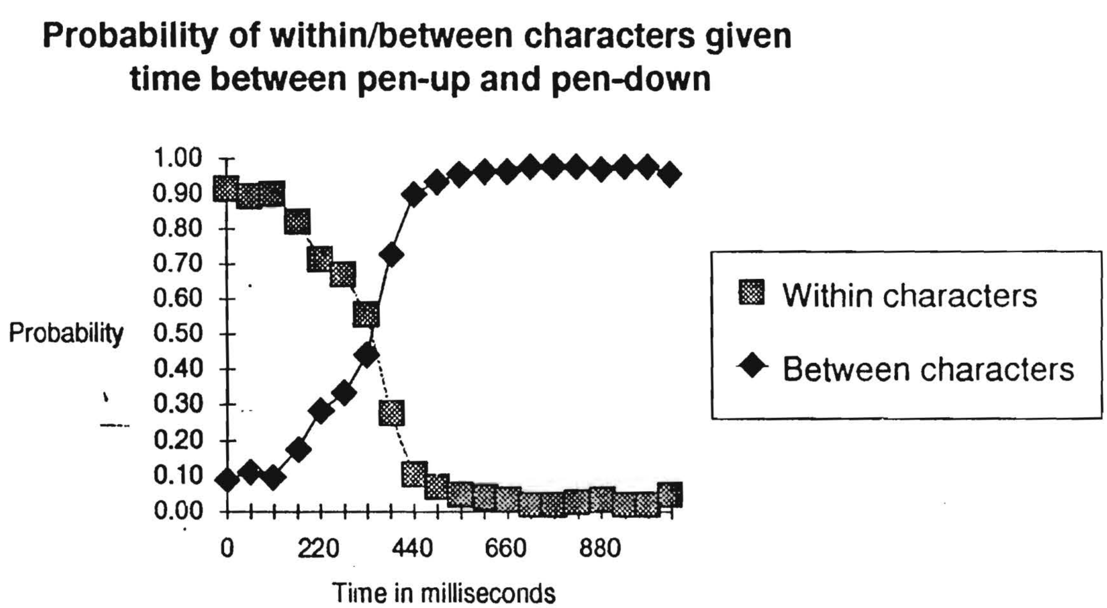

Figure 4

This chart shows that for pen-up times of 110 milliseconds or less, there is approximately a 90% probability that the strokes blong to the same character, and consequently a 10% probability that they belong to different characters. At 500 milliseconds or above, those probabilities are approximately inverted. Between 110 and 500 milliseconds, the probability curves cross in an approximately linear curve.

Pen-up velocity within and between characters

It may be argued that the results of the previous section (figures 3 and 4) are biased by the fact that strokes belonging to different characters are generally spaced further apart in space, and therefore it takes longer time just to move the stylus due to the increased distance. If this is true, then the pen-up time is of no relevance to the segmentation problem, because it adds no information to the spacial information.

In order to prove or disprove this fact, we analyzed the pen-up velocity. For each of the pen-up events considered in the previous section, we calculated the distance between the pen-up location and the immediately following pen-down location, and divided distance by time to obtain the velocity of the pen movement. Figure 5 plots the results, as obtained from the data collected in this experiment.

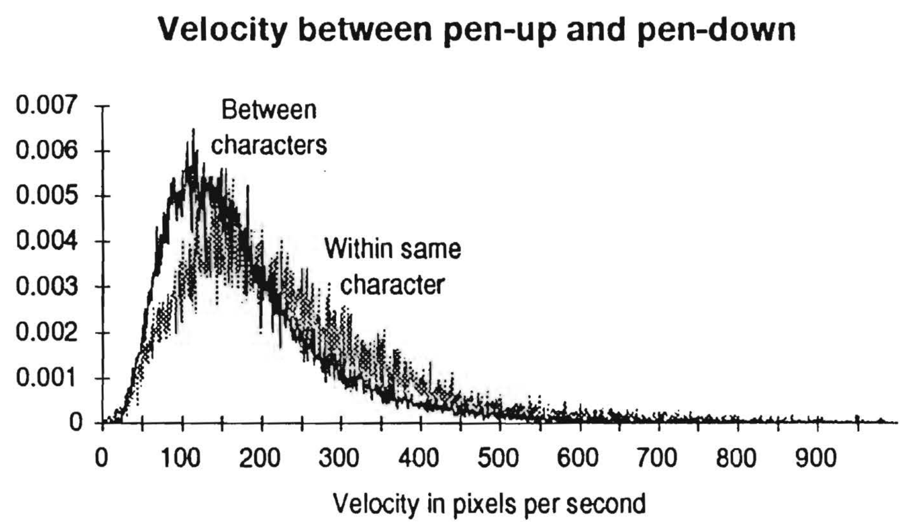

Figure 5

Then again, we asked: given a velocity in pixels per second, what is the probability that the two strokes belong to the same character. The data from this experiment is presented in a more appropriate form to answer this question in figure 6.

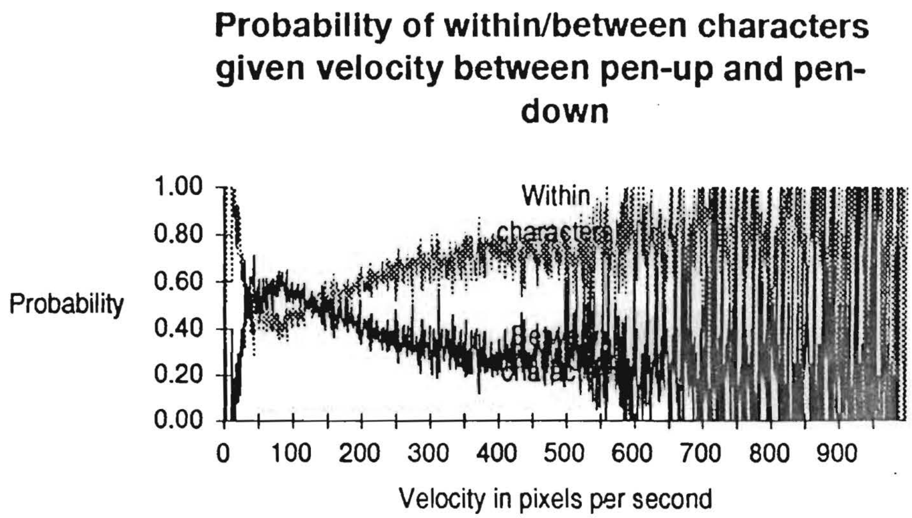

Figure 6

Figure 6 is shows that for low velocities (up to 30 pixels per second), there is a large probability that the strokes are within the same characters. Then, for intermediate velocities, the trend inverts and at around 100 pixels per seocnd there is a 60% probability that the strokes belong to different characters and about a 40% probability that they are part of the same character. Then the trend inverts once more, and for larger velocities (particularly for velocities greater than 300 pixels per second), there is a fairly constant 70% probability that the strokes belong to the same character.

Implications of the results

The data and analysis above suggest that stroke temporal information can be used to enhance the accuracy of the segmentation task in character recognition. Algorithms should look at character formation information based on spatial and temporal information. Then, they should attempt to recognize characters based on joint probabilities given by the visual properties of the characters and by the timing associated with their segmentation.