THE EFFECT OF A WORD PREDICTION FEATURE ON TEXT GENERATION RATE

Heidi Horstmann Koester and Simon P. Levine

Rehabilitation Engineering Program and Graduate Bioengineering Program
University of Michigan

Abstract
This study examines how use of a word prediction feature affects text generation rate performance. Fourteen subjects transcribed text with and without a word prediction feature for seven test sessions. Eight subjects were able-bodied and used mouthstick typing, while six subjects had high-level spinal cord injuries and used their usual method of keyboard access. Use of word prediction decreased text generation rate for the spinal cord injured subjects and only modestly enhanced it for the able-bodied subjects. This suggests that the cognitive cost of using this word prediction system largely offset the benefit of the keystroke savings achieved by these subjects.

Background
Word prediction is an effective way of reducing the number of selections required to generate text. This benefit in keystroke savings provides decreased motor requirements. However, it also exacts a cost in the additional cognitive and perceptual activities that are necessary to navigate the word prediction list [1,2].

The available data on user performance with word prediction suggests that the time required for these additional processes at least partially offsets the benefit of keystroke savings [3,4,5,6,7]. The keystroke savings reported for users of word prediction is fairly large, averaging 40%. However, the overall improvement in text generation rate for these users ranges widely; while some users enjoyed substantial improvement relative to letter-by-letter spelling, others improved only marginally or even decreased in rate. This provides indirect yet strong evidence that the cognitive time cost associated with word prediction can have a major impact on user performance. The wide range of reported improvements also suggests that this cognitive cost is highly variable, and the reasons for this variability need to be better understood.

Research Questions
The goal of this paper is to provide further insight into how the trade-off between decreased motor and increased cognitive loads affects text generation rate during use of a word prediction system. Performance will be measured under a range of usage conditions, to help determine the role of factors such as the characteristics of the system, the user, and the way in which the user employs the system. Ultimately we would like to define the conditions under which word prediction improves text generation rate and those under which it does not.

Methods
Subjects. Fourteen subjects participated. All shared the following characteristics: at least some college-level education; high familiarity with the standard keyboard; no significant prior experience with word prediction; and no cognitive, perceptual, or linguistic impairments. Eight of the subjects were able-bodied, while the remaining six had spinal cord injuries at levels ranging from C4 - C6.

Systems. The "Letters-only" system involved letter-by-letter spelling on a standard computer keyboard, and the "Letters+WP" system used single letter entry augmented by a word prediction feature. A six-word prediction list with a fixed word order was used, presented vertically in the top left corner of the screen. Able-bodied subjects used mouthstick typing, while subjects with spinal cord injuries used their usual method of keyboard access, which was mouthstick typing for two of the subjects and hand splint typing for the other four.

Experimental Design. An alternating treatments design was employed, in which subject performance with and without word prediction was recorded in each of seven test sessions. The keystroke savings provided by word prediction was fixed across Sessions 1 - 4 and varied in Sessions 5, 6, and 7. Each subject was assigned a particular strategy with which to use the word prediction feature. Labels for the four subject groups are shown below:

	SCI No	SCI Yes
Strategy 1	AB1 (n=4)	SCI1 (n=3)
Strategy 2	AB2 (n=4)	SCI2 (n=3)

Training. In the first part of training, subjects were instructed in the text transcription task and use of the mouthstick, for able-bodied subjects. Subjects then practiced using the Letters-only system for six blocks of text (four sentences each). The second part of training introduced subjects to the Letters+WP system and their assigned strategy for its use. The rule for Strategy 1 was to search the list before every selection. The rule for Strategy 2 was to choose the first two letters of a word without searching the list, then search the list before each subsequent selection. For both strategies, a search was not required when the list was empty. Subjects practiced using their strategy for four blocks of text (4 sentences each), which was sufficient for each to use the strategy correctly without prompting.

Testing. Each of the seven test sessions involved four sentences of warm-up using word prediction, an eight sentence test with word prediction, then a two sentence typing test. Text blocks were drawn from published typing tests [9] and revised to provide the levels of keystroke savings shown in Figure 1.

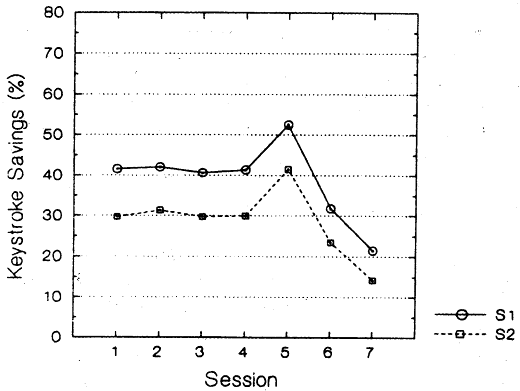

Fig. 1. Keystroke savings for strategies S1 and S2.

Sentences were presented singly on index cards. Subjects had twenty seconds to read the sentence before an audio cue signalled them to begin transcription, during which they could refer back to the card if necessary. Errors could be corrected by selecting the "Backspace" key as well as a special key for correcting word list selections.

Data Collection and Filtering. Subject behavior was recorded on videotape. Additionally, selected items were timed and stored by the software in real time. These data were filtered to remove events judged to be in any of the following three categories: text errors and error corrections; words not entered in a manner consistent with the assigned strategy; and "card reads", or times when the subject referred back to the text card during transcription.

User Performance Measures. Text generation rate and item selection rate were measured at each test session for both the Letters+WP and Letters-only systems. Text generation rate was defined as the number of characters generated during the test divided by the total time required to generate those characters. Item selection rate was defined as the number of items (i.e., keystrokes) selected during the test divided by the total time.

Statistical Analyses. Statistical differences in the dependent measures were determined using a repeated measures ANOVA technique. The between-subjects factors were strategy and presence/absence of spinal cord injury, and the within-subjects factors were system and session. Statistical significance for each ANOVA test was judged at a familywise p-value of 0.05.

Results

Filtering. The percentage of data removed from analysis averaged 16.3% of all Letters+WP selections and 7.3% of all Letters-only selections. The total amount of data filtered was independent of spinal cord injury, strategy used, or session.

Text Generation Rate. Subjects with spinal cord injuries averaged 116 characters/minute with Letters-only typing, which was significantly faster than the able-bodied subjects, who averaged 70 char/min ($p=0.005$). In contrast, subjects' text generation rates with the Letters+WP system were strikingly similar, averaging 71 char/min, with no statistical differences due to strategy or spinal cord injury.

The difference between spinal cord injured and able-bodied subjects re-emerged in examining the net change in text generation rate with Letters+WP relative to Letters-only, as shown in Figure 2 ($p < 0.0005$). For spinal cord injured subjects, word prediction had a strongly negative impact on text generation rate; on average, rate decreased by 40.7% when word prediction was used. For the able-bodied subjects, text generation rate was not significantly affected by the use of word prediction, except during Session 5, which had the highest level of keystroke savings and improved rate by 31.9%, and Session 7, which had the lowest keystroke savings and inhibited rate by 14.0%. Strategy of using Letters+WP had no effect on rate improvement for the able-bodied subjects, while spinal cord injured subjects who used Strategy 2 had a significant advantage over those who used Strategy 1 ($p=0.014$).

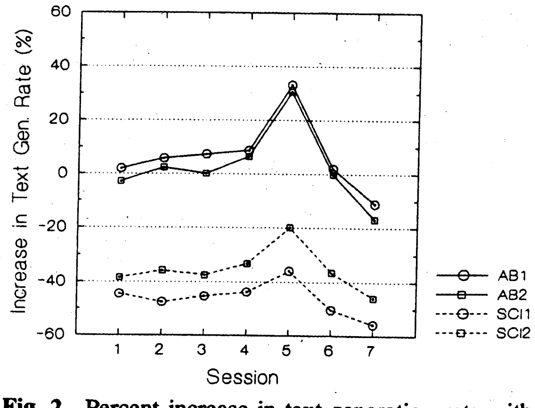

Fig. 2. Percent increase in text generation rate with Letters+WP, relative to Letters-only.

Item Selection Rate. For all subject groups, the item selection rate was significantly slower for the Letters+WP system than for Letters-only ($p < 0.0005$). Figure 3 illustrates this decrease as a relative percentage of the item selection rate with Letters-only. The effect of word prediction on item selection rate was larger for the spinal cord injured subjects ($p < 0.0005$); their item selection rate was

61.5% slower with Letters+WP as compared to Letters-only, while for able-bodied subjects, the average decrease in item selection rate was 31.8%. The strategy with which Letters+WP was used also influenced the decrease in item selection rate ($p < 0.003$). For both able-bodied and spinal cord injured subjects, item selection rate decreased less with word prediction for those who used Strategy 2 (which involved fewer list searches).

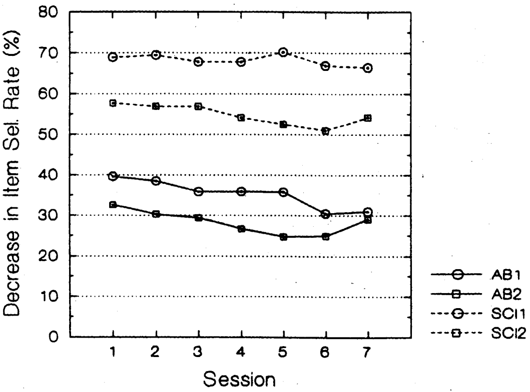

Fig. 3. Percent decrease in item selection rate with Letters+WP, relative to Letters-only.

Discussion

These results provide additional support for the hypothesis that increased cognitive and perceptual loads have a major impact on performance with word prediction. Any improvements in rate with word prediction relative to letters-only typing were much less than would be expected based on keystroke savings alone. Additionally, a statistically significant improvement was seen only for the able-bodied subjects, and only for the session that provided the highest keystroke savings. In all other sessions, able-bodied performance with word prediction was not significantly faster than without, while for spinal cord injured subjects, performance with Letters+WP was significantly worse than for Letters-only typing.

To determine the net effect of word prediction on performance, both keystroke savings and the cognitive cost of using the system must be considered. For example, use of Strategy 2 in this study provided lower keystroke savings but yielded performance at least as good as Strategy 1, because the fewer list searches required with Strategy 2 exacted a lower cognitive cost. As a second example, spinal cord injured and able-bodied subjects achieved the same keystroke savings, but the spinal cord injured subjects did much worse with word prediction than the able-bodied subjects, relative to letters-only typing. This suggests that the cost of word prediction was higher for the spinal cord injured subjects, which may be related to their greater *a priori* skill in letters-only typing.

The generalizability of these results is limited by features of the experimental conditions. Subjects were constrained in what strategy they were to use with Letters+WP, the text they were to generate, and the number of sessions in which they used the systems. Additionally, the spinal cord injured subjects represent only one sub-group of the actual user population, which includes individuals with more variable motor skills as well as those with cognitive impairments. Future work should focus on the performance of users with different abilities and levels of expertise than the subjects studied here, to either corroborate these results or reveal the conditions under which word prediction does provide a large improvement in rate.

References

1. Soede M, Foulds RA. (1986). Dilemma of prediction in communication aids. *Proc. of 9th RESNA Conf.*, Wash., D.C.: RESNA, 357-359.
2. Horstmann HM, Levine SP. (1991). The effectiveness of word prediction. *Proc. of 14th RESNA Conf.*, Wash., D.C.: RESNA, 100-102.
3. Koester HH, Levine SP. (in press). Learning and performance of able-bodied individuals using scanning systems with and without word prediction. *Assistive Technology*.
4. Newell AF, Booth L, Beattie W. (1991). Predictive text entry with PAL and children with learning difficulties. *British J of Educ Techn, 22:1*, 23-40.
5. Newell AF, Arnott JL, Waller A. (1992). On the validity of user-modeling in AAC: comments on Horstmann and Levine (1990). *AAC, 8:3*, 89-92.
6. Scull J, Hill L. (1989). A computerized communication message preparation program that 'learns' the user's vocabulary. *AAC, 4:1*, 40-44.
7. Venkatigiri HS. (1993). Efficiency of lexical prediction as a communication acceleration technique. *AAC, 9:3*, 161-167.
8. Koester HH, Levine SP. (1993). A model of performance cost versus benefit for augmentative communication systems. *Proc. of 15th IEEE-EMBS Conf.*, 1303-1304.
9. Lessenberry D. (1975). *College Typewriting*. Cincinnati: Southwestern Publishing Co.

Acknowledgements

Many thanks to Aram Sherman for his assistance in performing this study and to the subjects for their participation. This work was supported by the National Science Foundation, the University of Michigan Rackham School of Graduate Studies, and the U-M Rehabilitation Engineering Program.

Address

Heidi Koester
1C335 University Hospital
Ann Arbor MI 48109-0032
(313) 936-7170
Internet: hhk@umich.edu