QUANTITATIVE INDICATORS OF COGNITIVE LOAD DURING USE OF A WORD PREDICTION SYSTEM

Heidi Horstmann Koester and Simon P. Levine

Rehabilitation Engineering Program and Graduate Bioengineering Program
University of Michigan

Abstract

This study investigates the time cost associated with the cognitive processes performed during use of a word prediction system. Fourteen able-bodied and spinal cord injured subjects transcribed text with and without word prediction for seven test sessions. The extra time required to make word prediction selections, as well as times for keypress and list search actions, were measured as indicators of cognitive load. All subjects had slower keypress times during word prediction use, compared to letters-only typing, and spinal cord injured subjects had slower list search times than able-bodied subjects.

Background

The importance of cognitive load in determining user performance with word prediction systems is well-known. Task analyses have identified some of the component processes that may contribute to cognitive load, such as deciding when to search the prediction list and the visual search itself [1,2]. An important question continues to be how to quantify the effect and magnitude of this load. Measures of overall text generation rate provide a broad indication of the extent to which the cost in increased cognitive load counteracts the benefit of keystroke savings [3]. A more specific measure of cognitive load is extra selection time, defined as the additional time required to make each selection in a word prediction system relative to letter-by-letter typing. Extra selection time is assumed to be a direct reflection of the time spent on the additional cognitive processes.

Current knowledge about extra selection time is limited. A theoretical analysis has estimated extra selection time at 1.22 seconds, using results from information processing psychology [2]. The one reported empirical measurement is consistent with this, showing a range of 0.9 to 1.5 seconds, but measurements were made only for able-bodied individuals under a limited set of conditions [3].

Research Questions

This paper focuses on one aspect of an experiment in which the effects of a word prediction feature were studied with respect to overall text generation rate and cognitive load. While these two aspects are closely related, their differences warrant separate treatment. Specifically the goal of this study is to provide further understanding of the extra selection time associated with word prediction use, by measuring it under a broader set of usage conditions than have been studied previously. Additionally, we hope to measure cognitive load more precisely by separating the time for the primarily motor act of pressing a key from the primarily cognitive-perceptual act of searching the list. This is important in understanding the relative contribution of motor and cognitive factors to overall performance as well as a source of model parameters for future simulations of user performance.

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.

Interfaces. The "Letters-only" system involved letter-by-letter spelling on a standard computer keyboard, and the "Letters+WP" 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. Three spinal cord injured and four able-bodied subjects were assigned to use one of two word prediction strategies to form four subject groups: SCI1, SCI2, AB1, and AB2.

Training. In the first part of training, subjects were instructed in the text transcription task, and 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. All 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 [4] and revised to provide specific levels of keystroke savings. Sentences were presented singly on index cards which remained visible throughout transcription. Subjects had twenty seconds to read the sentence before an audio cue signalled them to begin transcription.

Data Collection. Subject behavior in each session was recorded on videotape. Additionally, all selected items were timed and stored by the software in real time. These raw 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, as identified through analysis of the videotape records.

Dependent Measures. Extra selection time was measured in each session and defined as the difference in selection times between the Letters+WP and Letters-only systems for that session. Selection time was defined as the total time required for a test divided by the number of selections (i.e., keystrokes) made during the test.

To measure cognitive load more precisely, the overall act of making a selection with Letters+WP was partitioned into the motor act of pressing the key and the cognitive-perceptual act of searching the word list. Each selection in a test was labelled according to whether it involved a keypress preceded by a list search or a keypress with no list search [5,6]. For example, when using Strategy 2, the first two letters of every word involved no list searches, so they were labelled as a keypress only. The keypress time (t_k) was then measured by averaging the times for all keypress-only selections. The list search time was calculated by subtracting one t_k from the time recorded for each list search-plus-keypress selection, then averaging the remaining times.

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 within each test was judged at a familywise p-value of 0.05.

Results

Extra Selection Time. Figure 1 shows the extra selection times for the four subject groups. Spinal cord injured subjects had significantly larger extra selection times than able-bodied subjects, averaging 0.910 seconds compared to 0.413 seconds ($p < 0.0005$). For all subjects, extra selection time decreased significantly as subjects gained experience with the Letters+WP system ($p < 0.0005$). Subjects who used Strategy 2 generally had lower extra selection times, since fewer list searches were required, but the difference was not statistically significant.

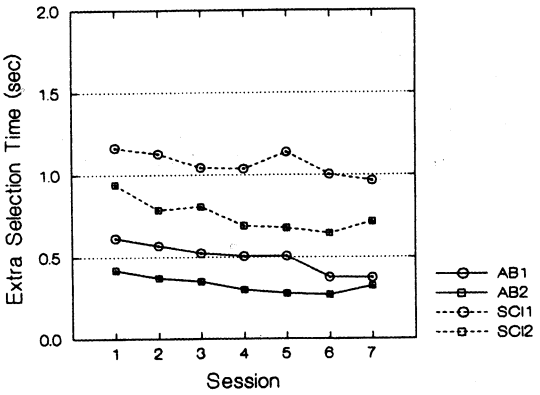

Fig. 1. Extra selection time with Letters+WP, relative to Letters-only.

Keypress-only Time. The average keypress times during use of Letters+WP are shown in Figure 2. There were no statistically significant differences between the groups, either on the basis of strategy or spinal cord injury. Session did have a significant effect ($p=0.001$), as keypress time improved an average of 17.7% from Session 1 to Session 7.

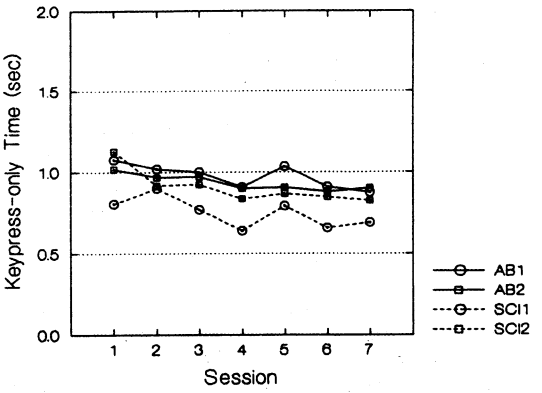

Fig 2. Keypress times during use of Letters+WP.

An unexpected result was that keypress times during use of Letters+WP were significantly slower than during Letters-only typing ($p < 0.0005$), with an average difference of 23% (170 msec). The keypress slow-down was more pronounced for subjects with spinal cord injuries than those without, with a slow-down of 48% (270 msec) for spinal cord injured subjects, and 10.8% (94 msec) for able-bodied subjects. However, this difference was not quite statistically significant.

List Search Time. Figure 3 shows the average list search times for the four subject groups. As with keypress time, strategy of use did not significantly affect list search time ($p=0.058$). Spinal cord injury, however, did have a significant effect ($p < 0.0005$); the list search times of subjects with SCI were an average of 96.4% (560 msec) slower than the able-bodied subjects. For able-bodied subjects, list search time improved an average of 27.3% (180 msec) from Session 1 to Session 7 (significant at $p < 0.0005$). For spinal cord injured subjects, however, list search time improved only 2.7% over these sessions, which was not significant ($p=0.395$).

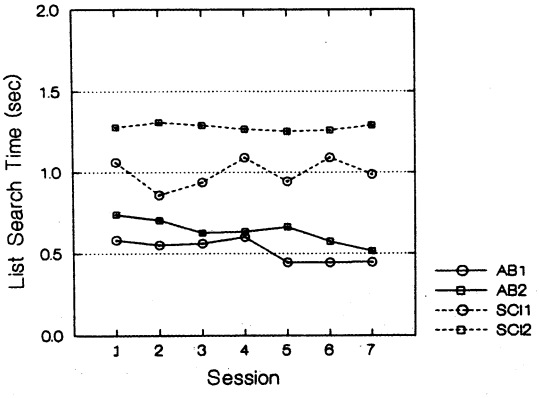

Fig 3. List search times during use of Letters+WP.

Discussion

These results suggest that using word prediction exacted a substantial cognitive cost for these subjects. Based on the extra selection time results, the execution of cognitive and perceptual processes comprised between 30% and 60% of the time spent using the Letters+WP system. The sources of this cognitive load and possible reasons for the differences seen between subject groups are discussed below.

List search time was a major contributor to cognitive load, measuring at least several hundred milliseconds for all subjects. However, it is unclear why the list search time for spinal cord injured subjects was so much larger than for able-bodied subjects. This may have been due to a difference in searching styles; perhaps the spinal cord injured subjects performed more thorough searches, while the able-bodied subjects used anticipation to skim the list on occasion. A second possibility stems from the fact that the technique used to measure list search time also captured the time for other processes, such as deciding whether to search or verifying accuracy of a selection, if and when they occurred. These processes contribute to the cognitive overhead involved in monitoring and guiding overall activity, and this overhead may have been greater for spinal cord injured subjects. This would not be that surprising, since they had much more prior experience under the Letters-only condition than the able-bodied subjects, which would increase the cognitive difficulty of switching to the Letters+WP system.

Cognitive overhead may also account for the result that keypress time was slower with the Letters+WP system. If keypress time had measured only the motor component of item selection, as intended, its duration should have been essentially the same with and without word prediction. That it was not suggests that cognitive overhead was present even during selections that did not involve a list search [2]. The relatively large keypress slow-down seen for spinal cord injured subjects provides further evidence of greater cognitive overhead for these subjects.

Certainly future work is necessary to address these unresolved issues and to examine cognitive load under different sets of conditions. This study provides a methodological starting point for future work as well as intriguing initial results. Continued progress toward understanding cognitive processes and their associated time costs is critical in order to clarify how cognitive load impacts overall user performance and, ultimately, to determine ways of reducing its effects.

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. Lessenberry D. (1975). *College Typewriting.* Cincinnati: Southwestern Pub. Co.
5. Card S, Moran T, Newell A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Erlbaum Associates.
6. Olson JS, Nilsen E. (1988). Analysis of the cognition involved in spreadsheet software interaction. *Human-Computer Interaction, 3:4,* 309-350.

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