Measuring the effectiveness of word prediction: The advantage of long-term use

Tina Magnuson and Sheri Hunnicutt

Department of Speech, Music and Hearing, Stockholm, Sweden

Abstract
This paper provides results regarding keystroke savings and text composition time from a long-term evaluation of a Swedish word prediction program. The evaluation was carried out as a single-case pilot study for one participant with dyslexia and no motor disabilities. During 13 months of training, during which a speech synthesis facility was introduced, his performance was shown to improve both with and without the word prediction program after an initial increase in text composition time. On average, he was able to save 22.7% total keystrokes with word prediction only and 30.5% with word prediction combined with speech synthesis. Text composition time increased upon introduction of the word prediction program by at least 22 %. Over time, however, and with the further introduction of speech synthesis, he was eventually able to return to his original text entry speed.

Introduction

Predictive word processing systems have existed since the early 1980s and were originally intended as a speed enhancement support for persons with motor disabilities. Research has shown, however, that, contrary to expectations, word prediction can have the effect of slowing down the writing process (Horstmann Koester & Levine, 1994a; Venkatagiri, 1993), at least in the short term. Most researchers, though, are familiar with cases in which there has been substantial rate enhancement for some users (Newell, Arnott, Booth et al., 1992; Damsby, S-LP, personal communication). Characteristics of predictive systems that are more agreed upon are the distinct advantage of word prediction in effort savings due to fewer keystrokes being required for a given length of text, and the production of text that is more error free (see, for example, the study by Waller, Beattie, & Newell, (1991); with Predictive Adaptive Lexicon (PAL)). Our own research with the programs Predict (Hunnicutt, 1986) and its successor Prophet in the late 1980's and 1990's has confirmed these results.

Predict and Prophet were evaluated for several years, initially together with individuals with slow and laborious writing stemming from a motor dysfunction (Magnuson, 1994). As slow writing is often believed to be an important issue for individuals with motor disabilities, the program's original purpose was to accelerate the writing process for these individuals. Briefly, the results of these first studies indicated that:

There was most often a reduction of keystrokes, which meant less effort;

A reduction in the number of keystrokes did not necessarily lead to a savings in time;

The writing strategy had to be changed due to the possibly higher cognitive load on the writing process. In particular, the time saving gained by fewer keystrokes was found to be consumed by the longer time needed to search for the right alternative, which involved shifting one's eye gaze from the keyboard to the screen and back to the keyboard before making a decision and pressing the right key;

In contrast to what was expected, rate enhancement was not the most important aspect to users, but rather the effort saving (as typing often is very laborious for a person with a motor disability) and the possibility of producing more correct texts;

Objective measures (such as reduced number of keystrokes, time savings or improved spelling) did not always correspond to the participants' subjective experiences and preferences.

The participants reported positive subjective opinions such as "Prophet helps me to write

Speech, Music and Hearing, KTH, Stockholm, Sweden
TMH-QPSR Volume 43: 57-67, 2002.

more independently" and "I get less exhausted when I write with Prophet."

However, the interpretation of such findings has proven to be a complex issue, already addressed by numerous investigators. An early study by Soede & Foulds, (1986) discussed the problematic issue of the possible time savings resulting from a reduced number of keystrokes being consumed by a heavier mental load, a writing strategy change, and more time-consuming eye movement patterns. On the other hand, they conjectured that the possible negative effects of the cognitive load, including choice making, may be more than compensated by having the words visually presented and already correctly spelled. Moreover, Tyvand, Endestad, Pedersen and Heim (1994), found no significant slowing down due to eye movement pattern shifts, contrary to the findings in the study by Venkatagiri (1993), in which 10 out of 16 able-bodied writers took a longer time to produce sentences in a communication aid with predictions than without.

The issue of the effect of training and long-term use of word prediction on time savings has been addressed by Newell, Arnott, Booth et al., (1992). In this article, it was observed that rate enhancement and increased text production can be attained with predictive techniques for motorically disabled children and that improvements in spelling and sentence construction could be observed for persons with language dysfunction. However, text composition time and keystroke savings were not detailed. It was commented that "even such apparently objective measures as keysaving and rate of composition need to be verified by long-term case studies." The work described herein is part of such a study.

The present study

In order to further explore the relationship between these results and the findings from our earlier studies, we followed the long-term use of predictions by a user with dyslexia and previously limited use of computers. The primary purpose of the study was to investigate the qualitative effects that Prophet II (a successor to Prophet), based on a probabilistic (Markov) model, would have on this writer's linguistic expression. An earlier phase of the study showed that Prophet II supported spelling and morphology in a positive way. These results, together with a study exploring subjective experiences and readers' attitudes toward the quality of the texts, have been reported in a paper by Magnuson & Hunnicutt, (2000). A secondary purpose was to investigate the effectiveness of text composition using Prophet II.

Our goal in this part of the study was to determine whether extensive program use could reverse the expected effect of reduced writing speed. Our specific questions were

- to what extent the program could reduce the total number of keystrokes,
- to what extent the program would have an effect on text composition time,
- whether there was a dependent relationship between keystroke savings and text composition time and
- whether long-term use would have any effect on these factors.

We hypothesized that Prophet II could reduce the number of keystrokes needed for a certain amount of text, since we had seen this in the earlier studies (ours as well as others); and that there would be no immediate time savings since many studies have reported that word prediction *slows down* the writing process rather than enhancing speed; if these first two hypotheses hold we would be able to confirm that fewer keystrokes do not automatically result in a reduced amount of text composition time when using word prediction. However, we also hypothesized that long term usage and becoming more familiar with the new text entering strategies would be beneficial in text composition time.

Method

Equipment

In order to make predicted words grammatically probable, a probabilistic language model was employed in Prophet II (Hunnicutt & Carlberger, 2001). Two Markov models are included, one for words (word unigrams and bigrams) and one for word classes (word class unigrams, bigrams and trigrams). A tagged text corpus was used to train a tagger (Ejerhed, Källgren, Wennstedt, & Åström, 1992) utilizing about 150 word class tags. The level of detail of these tags is illustrated by nouns, which are subdivided according to gender, plurality, definiteness, and sometimes case. A limited set of word endings was compiled and used to give information about the probability distribution of word classes of new words as they are learned by the program. Additional keystroke savings is provided to the program by a number of heuristics including recency promotion, the choice of not repeating a prediction until all other predictions have been presented, sensitivity to upper and lower case letters and the possibility to use up to 9 predictions. A simulation with unigram and bigram word lexicons of about 10,000 entries each resulted in a simulated keystroke savings of 46% with five predictions.[1]

Also during the present study, a list consisting of five predictions was used. The target word could be selected through any of the following alternatives: a) by clicking the target word in the list, b) by highlighting the target word in the list and pressing the return key or, c) by using the function keys, which corresponded to the predicted words. The possibility of adding new words to the database after a writing session was not exercised.

A feature of the program is a log file that was designed for evaluation and research purposes. The log file registers keyboard and mouse operations. The information that can be extracted from the log file includes aspects such as deletions, arrow movements, pause time, and acceptance of a predicted word. The data could be imported into Microsoft Excel for further analysis.

Speech synthesis

A speech synthesis system, the Infovox 220, was made available. By means of different settings, the speech synthesis could read out lists of suggested predictions, menu alternatives, entered words and highlighted text.

Participant

A 25-year-old adult male with Swedish as his first language, who has dyslexia but no physical impairment participated in the study. His difficulties included severe spelling problems, somewhat milder, but still severe, reading problems (at the 15th percentile for grade 8 on a word decoding test) and no other immediate overt manifestations of linguistic dysfunction. He reported word-finding obstacles, especially when tired, which was supported by the result of a word retrieval test (Boston Naming Test, Swedish version). These difficulties could be averted to some extent with aural prompting. He was also uncertain of the pronunciation of long words, a phonological deficit that was not conspicuous in normal, relaxed conversation. He was not what is called a reluctant writer, and he had many stories to tell, but his writing was mostly private due to his literacy problems. It was not feasible to produce a more detailed, linguistic profile as the participant earlier had gone through demanding and extensive testing experiences and was reluctant to be tested further. This person had himself taken the initiative to complete his high school education and worked very conscientiously to obtain good results. He was motivated to try out new techniques and aids although his experience with computers was very limited. At school he had mostly been using pen and paper. He had a previous vocational education, but even that professional situation, because of its pace and stress, was too demanding of his reading abilities.

Study design

The study was designed to measure possible keystroke and time savings in texts written in Swedish with predictive writing support in order to provide answers to our research questions. Since we wanted to evaluate this particular writer's performance, a single-case design was applied, and in this pilot case, an alternating treatment study was chosen. His writing experience was limited, so we expected a general improvement of his writing skills regardless of writing condition. For this reason, we did not choose the usual response guided A-B-A design, rather the *without* and *with* prediction conditions were planned to be exactly alternating in order to detect a difference between conditions. The trends within each condition could thus be compared without confusion by a general improvement over time or between sessions. Two weeks prior to the first measured writing session (session 1) the participant borrowed a computer, with Microsoft Word installed, from our department. The pre-study phase and baseline condition consisted of text production using only the word processor, the without Prediction condition (w/o P). The word prediction program (Prophet II) was introduced at the second measured writing

[1] For further information about the improved version of Prophet, see Carlberger & Hunnicutt (1997); Carlberger (1997); Magnuson & Carlberger (1997)).

session and text production with its use gave rise to the Word Prediction condition (WP).

When the writer, after six sessions, indicated that he was now ready to use the speech synthesis option, the prediction feature was combined with a speech synthesis module, the Word Prediction + Speech Synthesis condition (WP+SS). He used it mostly to listen to suggestions in the prediction list, and to check entered words. The participant was able to use the equipment as much as he wanted between sessions, but no specific training tasks were required. He reported that he used the program mostly for schoolwork.

The test material consisted of texts that were produced during thirteen sessions in a free writing task about a self-selected topic for a given time, 40 minutes. Thirteen texts were thus collected at thirteen separate sessions during a period of thirteen months, the sessions taking place at his school or home or at our department. However, the planned alternating order was not strictly adhered to. Rather, according to the writer's wishes, the order was sometimes changed, resulting in a less preferable allocation that was not randomized or preset but rather, was opportunistic. At the time, however, this change was judged to be of minor importance. The arrangement of the alternating conditions can be seen in Figure 1. These sessions thus contained the alternating conditions; a) a word processor (the w/o P condition), b) a word processor with word prediction (the WP condition), and c) a word processor with word prediction and speech synthesis (the WP+SS condition).

Keystroke definition and registration

Registration of text composition time and keystrokes was made possible by means of the log file previously mentioned. The definition of keystroke savings used is an adaptation of a common one for direct selection: the percentage of keystrokes eliminated (saved) in the test texts by employing word prediction with standard keyboard input instead of standard keyboard input alone. This measure is chosen, not because it gives a total picture of the help provided a user, but because it is a simple measurement which gives a commonly employed and well understood measure of the success of the algorithm itself, and of its associated lexicons, in making appropriate predictions. The adaptation was made as follows: the ratio of total keystrokes to the number of characters thereby produced in the final text was calculated (instead of percentage); mouse clicks for selecting a word were registered as one keystroke; mouse movements, for the purpose of highlighting and/or listening, were not counted as keystrokes. The keystroke registration allowed for a comparison of the total number of keystrokes, first with entered characters and then with the characters visible in the final text. The total keystrokes included *indirect* operations such as carriage return, arrow movements and operations leading to deletions, as well as *direct* operations (the entered characters) which resulted in a letter, space, number or predicted word visible in the text, i.e.,

final text = total # of kbd and mouse ops – indirect ops – deleted direct ops

where kbd = keyboard and ops = operations. The final text, then, is composed of the text resulting from the undeleted direct operations.

Results and discussion

Because the user interface for the program was still under development, some of the functionality we needed for our evaluation was not stable. The logging facility malfunctioned four times with the consequence that no keystroke registration occurred for two texts without prediction (texts number 6 and 8) and for two texts with prediction (texts number 7 and 9). These texts are hence excluded from all calculations regarding keystroke savings and the

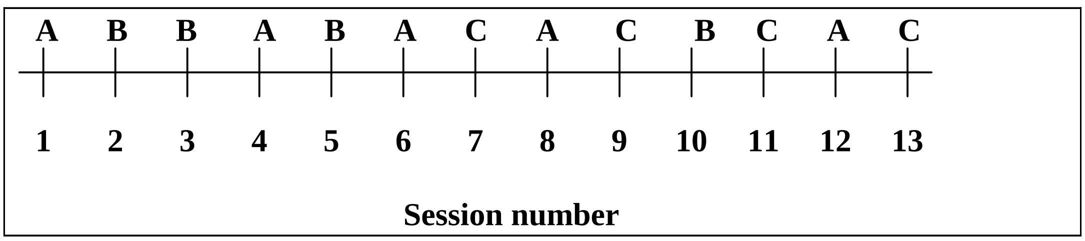

Figure 1. Arrangement of the alternating writing conditions. A = the w/o P condition; B = the WP condition: C = the WP+SS condition.

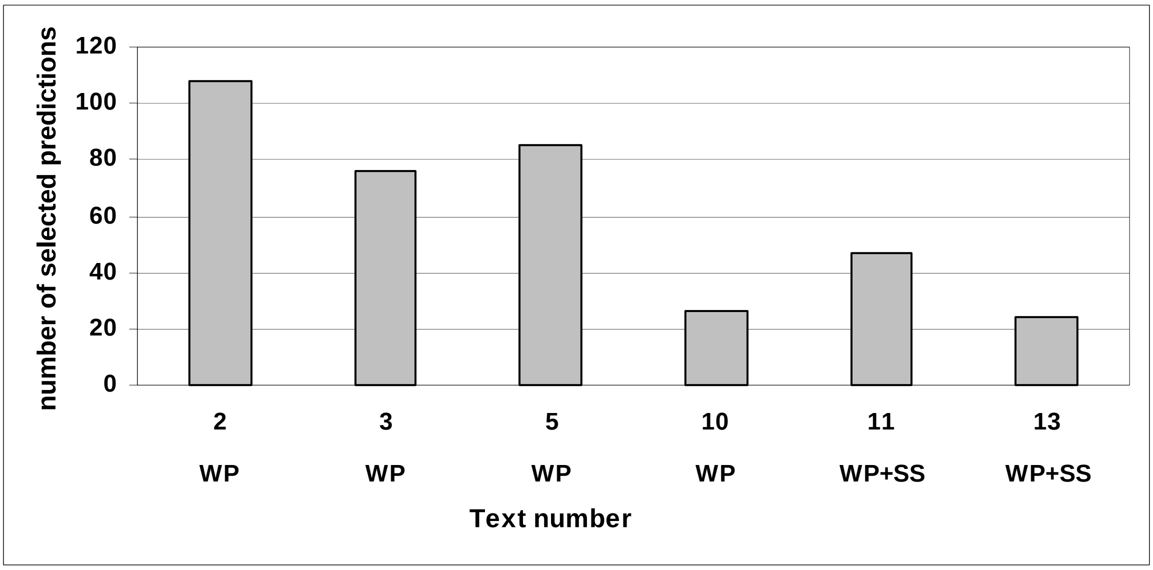

Figure 2. Texts written with predictions. The number of selected predictions per text. Texts numbers 7 and 9 are excluded.

relationship between time savings and keystroke savings. The time registration for text 7 was found to be unreliable as well, so this text is excluded from the time savings calculation.

As previously mentioned, the participant was asked to write on a self-selected topic for 40 minutes. However, as the experimenter determined that it was more practical to allow for the completion of an initiated sentence or paragraph, or for somewhat early termination due to fatigue or lack of further ideas, the actual task time varied from 30 to 47 minutes. Text composition time was measured with a stopwatch and verified with the time registration in the log files. For texts written under the condition without prediction (the w/o P condition), the length varied from 225 to 379 words (mean 288 words, 42.8 minutes), whereas texts written with predictions but without the aid of speech synthesis (the WP condition) varied from as few as 144 to 223 words (mean 179 words, 39.6 minutes) and texts under the WP+SS condition (prediction with speech synthesis) from 195 to 288 words (mean 249 words, 31.6 minutes). This decrease in text length under the WP condition may be an indication of the participant's decoding difficulties and the novelty of the technique or that the creative writing process was interrupted by appearance of the predictions on the screen. In the prediction mode, predicted words were selected between 24 and 108 times (mean 61 selection) per text. The number of selections per text is shown in Figure 2. We can observe that this value decreased as the study proceeded. Inspection of the log file shows that the participant deleted more of the accepted predictions at the beginning of the study and kept the accepted predictions in the later texts. One plausible explanation is that, when supported by the speech synthesis (sessions 11 and 13), he was able to aurally check the prediction alternatives and to aurally monitor what he had already written, and thus may have begun to make more careful selections. An additional explanation can be that the accepted predictions in the first texts consisted mainly of function words of three letters or fewer, whereas at the end, most of the accepted predictions were content words, indicating that, as time went on, he chose to type the short and numerous function words that he knew how to spell and to use the predictions primarily for longer content words. This explanation seems to be contradicted by text 10, in which speech synthesis was not available but few predictions were selected. In the section about time savings (Figure 5), however, we can see that the keystroke savings ratio was the highest for text 10, indicating that the prediction facility was rarely used here.

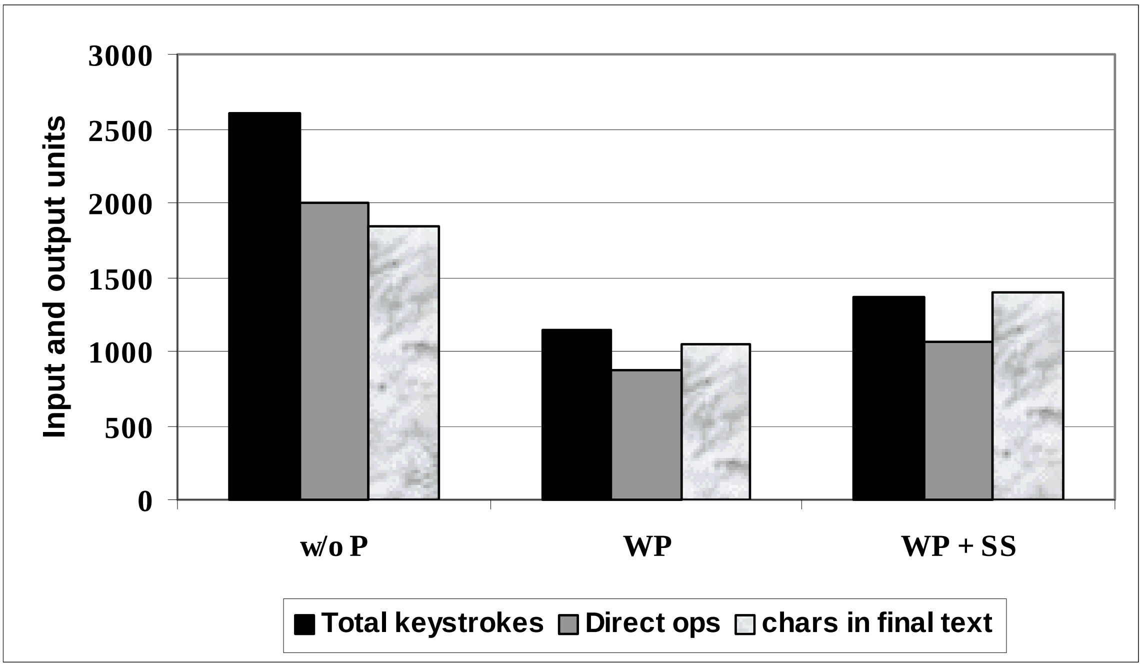

Figure 3. The number of total keystrokes compared to the number of direct operations only and the number of characters in the final text.

Keystroke savings

A keystroke savings ratio was calculated in Microsoft Excel from the log file in which time and type had been registered for each keystroke. The relationship between the total number of keystrokes and the number of characters in the final text is shown in Figure 3. The bars reflect the mean calculated without the results from texts 6, 7, 8 and 9 for which information was missing. *Total keystrokes*, as previously related, refers to the total number of keyboard and mouse operations: letters, numbers, space, punctuation marks, function keys for word selection, shift, arrow movements, deletions, carriage return, and mouse clicks for word selection and cursor positioning (direct and indirect operations); *direct operations* refer to keystrokes producing letters, numbers, spaces, and selection of predicted words, and *chars in final text* to the letters, numbers, punctuation marks and spaces in the final text (information given by the Word Count feature in Microsoft Word). As can be seen in Figure 3, there was a reduction of total keystrokes for the texts written with predictions, the ratio being 1:1.41 for the w/o P condition, 1:1.09 for the WP condition, and 1:0.98 for the WP+SS condition. This means that a sample text of 200 characters required 282 total keystrokes (direct + indirect operations) in the w/o P condition, 218 total keystrokes in the WP condition and 196 keystrokes in the WP+SS condition. In both the w/o P and WP conditions the number of total keystrokes exceeded the number of characters in the text. This finding may seem unexpected but is due to the fact that letters and numbers are accompanied by the indirect operations of deletions, arrow movements and other editing commands. In the two prediction conditions these indirect operations consumed part of the savings. The number of keystrokes relating to letters, numbers and spaces alone (direct operations) is, as expected, less than the number of characters in the final text in both prediction conditions, the ratio being 1:1.1 for the w/o P condition, 1:0.83 for the WP condition, and 1:0.76 for the WP+SS condition. This means that the same final text of 200 characters needed 220 direct operations to produce all letters and numbers in the w/o P condition but only 166 in the WP and 152 in the WP+SS condition.

In Figure 4 we can see that the keystroke savings ratio was kept approximately constant within the w/o P and the WP conditions, with the exception of text number 10. Performance improved in the second text representing the WP+SS condition with a keystroke savings ratio of 0.7 in text 13.

Between 22.7 % (WP) and 30.5 % (WP+SS) keystroke savings were achieved, which answers

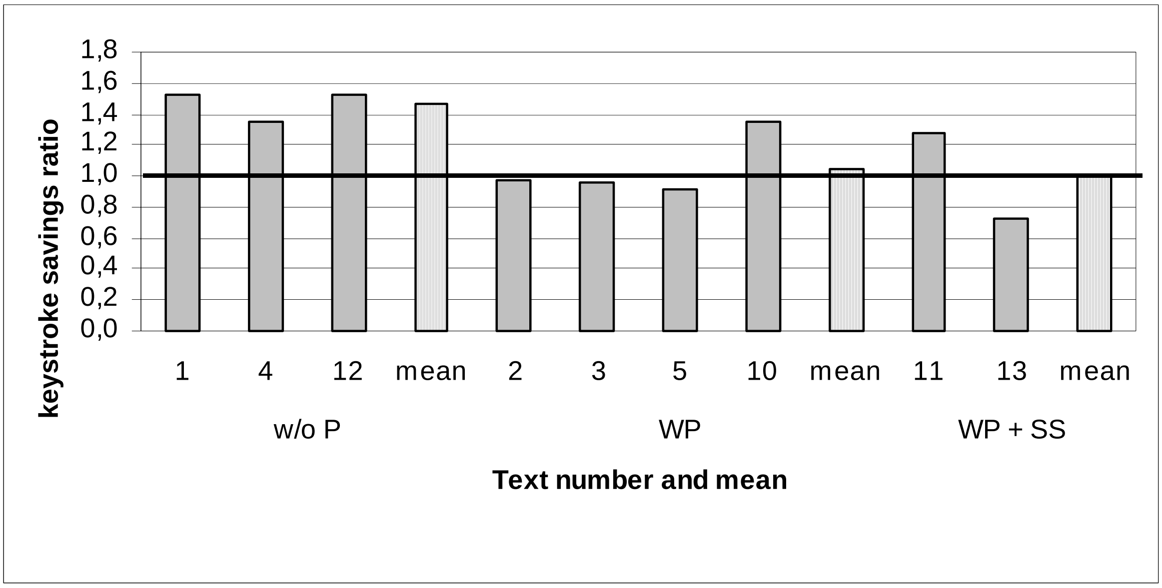

Figure 4. Changes in keystroke savings ratio over time and within condition. The horizontal line indicates the point at which one keystroke corresponds to one character in the final text.

the first question as to what extent the program could reduce the number of keystrokes (direct and indirect operations). Even though deletions and arrow movements consumed part of the savings, there was evidence of "real" savings (where the number of total keystrokes is less than the number of remaining characters in the final text, and thus the ratio between keystroke savings and characters in the final text is less than one) in the WP+SS condition. Additionally, the present study showed that the number of direct operations needed is smaller using word prediction, which agrees with our earlier studies (Magnuson, 1994).

Time savings

Time savings was calculated on writing speed, expressed in *words per minute* (wpm), i.e., the number of words in the final text divided by the number of minutes it took to produce the text. As is evident from Figure 5, there was initially a time loss rather than time savings, which is a result consistent with our earlier findings as well as those of other investigators (Horstmann Koester & Levine, 1994a; Venkatagiri, 1993). The increase in text composition time upon introduction of word prediction was 22% from text 1 (w/o P, 4.9 words/minute) to text 2 (WP, 3.8 words/minute) with an additional 4 % to text 3 (WP, 3.6 words/minute). Further text composition with word prediction alone consistently required more time than unaided composition. However, as this participant began to use Prophet II combined with speech synthesis, this result ceased to hold: by the end of the study there was no time loss in the WP+SS condition. As we can see from the trend-lines, the participant's writing rate improves within all conditions, but less so in the WP only condition. Since many of the basic strategies involved in using predictions, such as shift of eye gaze and lexical decision-making would be the same with or without speech synthesis, we can speculate that his severe decoding difficulties, and not the prediction strategy per se, would cause his slower text generation rate under the WP condition.

With regard to the aspect of writing rate, we noticed an inconsistency in performance from session to session, especially in the w/o P and the WP+SS conditions. Though large variability is a general observation for free writing, this exceptionally highly fluctuating performance among persons with writing difficulties has been observed and reported earlier (Laine & Follansbee, 1994). However, the issue of time savings has quite a different importance depending upon whether we study a group consisting of able-bodied persons or writers with physical disabilities. Although the time aspect may not be experienced as the most essential one for writers with motor disabilities (i.e., very slow writers) and/or persons with laborious writing for other reasons, the actual gain from word prediction is more substantial in this group. The slower the writing (e.g., for those

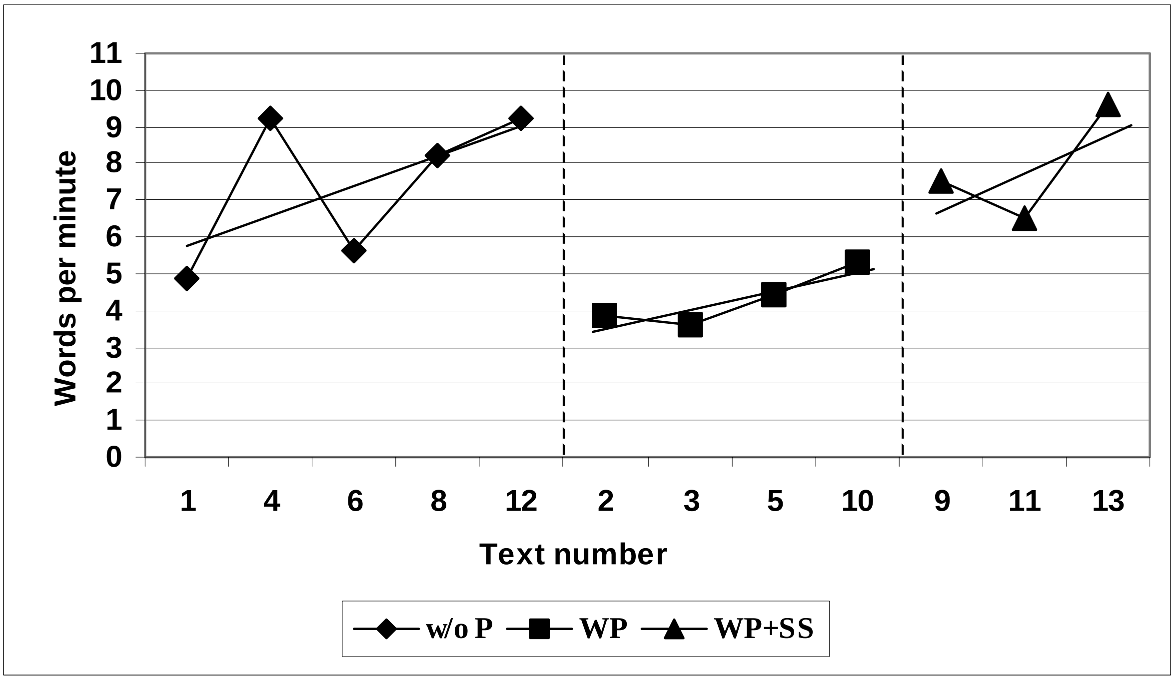

Figure 5. Writing rate expressed in words per minute. A trend line is inserted for each condition.

using scanning, a head stick or mouth stick), the more the time savings. However, as writing speed increases, there is a point at which the prediction condition ceases to improve writing rate. Newell, Arnott and Waller (1992, p. 91) propose that a possible estimation of reasonable time savings for very slow writers (less than 5 words per minute) is that the "increase in speed is approximately equal to the keysaving." In their example a 50% reduction in keystrokes needed corresponded to an increase of 50% in speed. In summary, earlier studies indicated that persons without any motor disability, even if they are slow writers due to factors such as dyslexia or lack of experience, do not write faster with prediction and that prediction techniques actually produce slower writing for these persons than ordinary word processing. With this study we have been able to indicate that speed can be enhanced within the WP condition through familiarity with the prediction technique, especially when combined with speech synthesis, even for persons with original typing speed faster than 5 words/minute.

The relationship between keystroke savings and time savings

We can thus concur with previous results that a reduction of keystrokes does not *automatically* lead to time savings. However, an interesting relationship for the current study can be observed on the scatter plot in Figure 6. If the number of words/minute and keystroke savings ratio were correlated, the expected result would be a linear function with negative slope, showing, for a given text, more words/minute with fewer keystrokes (low keystroke savings ratio) and fewer words/minute with more keystrokes (high keystroke savings ratio).

If we consider the means in the three conditions, one would then expect the means for the two conditions with word prediction to lie higher and to the left of the mean for the without prediction condition. That is, one would expect the two word prediction conditions to result in more words/minute with fewer keystrokes. As we can see from the figure, keystroke savings due to word prediction is not accompanied by an increase in words/minute. The condition with the best keystroke savings (ratio = 0.98 as compared to 1.41 in the w/o P condition) is the WP+SS condition, resulting in little, possibly insignificant, rate enhancement. There are also keystroke savings in the WP condition (ratio = 1.09), but the text composition rate (4.3 words/min) is substantially lower than in the other two conditions. It is interesting to note, however, that within each of the three conditions, the reduction in keystrokes often leads to time savings. This relationship would possibly imply a consistent user strategy within conditions.

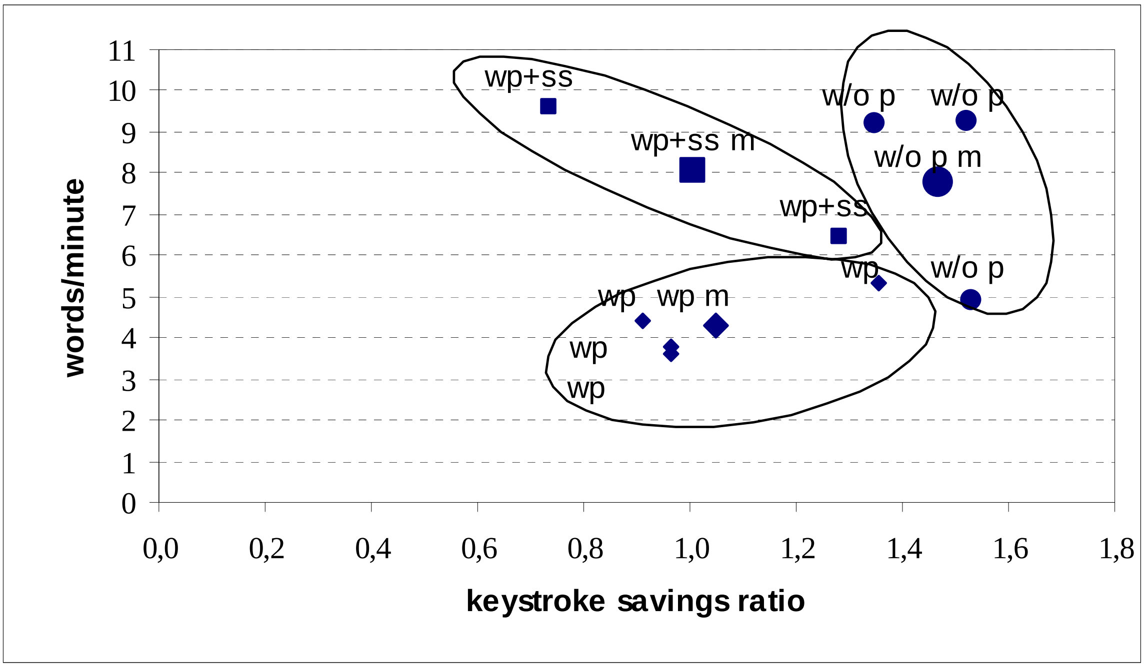

Figure 6. The relationship between keystroke savings and time savings. The larger data point symbols represent the mean (m) for the condition.

As previously mentioned, using prediction techniques involves a strategy shift and a heavier cognitive load, as well as a word decoding task, which implies more time for each keystroke. For persons with a motor dysfunction and very slow typing, the reduced number of keystrokes, despite the slower prediction strategy, compensates for the length of time consumed typing letter by letter. This is supported by the results from the study by Horstmann Koester & Levine, (1994a) in which able-bodied individuals (AB) using mouth-sticks were compared to persons with spinal cord injury (SCI) using their usual means of text entry. The AB group was much slower with the letter-by-letter typing strategy whereas the two groups showed no significant difference in the word prediction condition.

However, again we want to highlight that it is not clear whether time savings is the most important and desired feature of writing support. For many persons, such as those with laborious and tiresome writing or spelling difficulties, other aspects, such as less frustration and fatigue, have been shown to be equally important.

The effect of long-term usage

The results on possible time savings from the current study thus contradict earlier findings with individuals without motor difficulties, which have shown that the motorically non-disabled user is likely to write at a considerably lower speed using predictions. This participant returned to his original text entry speed after thirteen months of usage and, perhaps due to the decoding difficulties that accompany dyslexia, only when word prediction was combined with speech synthesis. He, like most other users, showed a reduction of keystrokes, but also a time loss, for the first three texts written with predictions. If the study had ended after five months we would not have been able to detect the improvement of text entry speed, which was equal to or slightly faster than ordinary word processing.

General discussion

An aspect that is sometimes overlooked is the importance of decreased motor effort. Horstmann Koester & Levine, (1994a, p. 117) state that their study "... provide[d] indirect yet strong evidence that the additional cognitive and perceptual activities reduce the benefit of decreased motor requirements." It seems, however, that decreased motor effort has a value in and of itself. The keystroke savings may be more crucial than time savings for an easily exhausted writer with motor dysfunction, which is also argued by Waller et al., (1991) and Newell, Arnott, & Waller, (1992).

A possible comparison with other studies is complicated by the fact that most evaluation studies of prediction techniques have been carried out under different conditions. Some studies have been conducted with able-bodied subjects (Higginbotham, 1992; Venkatagiri, 1993), others with able-bodied subjects compared to fast motorically affected writers ("spinal cord injuries at levels ranging from C4-C6") (Horstmann Koester & Levine, 1994a, 1994b) and still others as laboratory simulations (Palazuelos Cagigas, Godino Llorente, & Aguilera Navarro, 1997). Attempts have been made to design a standardized system assessment by referring to the "current international efforts on standardization of evaluation and assessment processes" (Palazuelos-Cagigas, Aguilera-Navarro, Rodrigo-Mateos, Godino-Llorente, & Martín-Sánchez, 1999). This standardized method can only constitute a first step towards a complete evaluation process, since it does not include individual user performance, which is also recognized by the authors.

Further, the whole issue of keystroke savings versus time savings is complicated by the most essential issue, namely the personal preferences of the individual writers. In Horstmann Koester & Levine, (1996) cost benefit model, the net gain is calculated as a cost, defined as the percent decrease in item selection rate divided by the benefit in the percent of keystrokes saved. However, the real benefit for persons with motor and/or language dysfunction differs from and is often much more complex than the cost benefit in the straightforward, unproblematic performance of an able-bodied writer in a laboratory setting. In our clinical experience we have met with extreme opinions about the potential benefit. One participant, for whom we could observe no literacy improvement nor time savings, could not manage to write without the program; another participant preferred her extremely slow typing even though we were able to show considerable improvement in speed. The point is that more qualitative data must be included in a functional evaluation of a prosthesis, something that also has been addressed by, for example, Newell, et al., (1992). Studies such as the one by Laine & Follansbee, (1994) with a group of low-functioning students with hearing impairment show that aspects such as disruptive behavior, focusing on the task, and the degree of written word fluency can be affected by the writing environment and prove to be of considerable importance to diverse disability groups. The various effects caused by prediction techniques, such as keystroke savings, time savings, a shift of cognitive focus, and linguistic support have to be weighted together with the personal profile and preferences of the writer to determine the net gain and to define the guidelines for writing aid selection.

Recommendations for future research

We believe that this method of a long-term, single-case study is an informative way to further investigate the advantages of prediction techniques and other writing supports. Due to the previously discussed shortcomings in the present study, this method could only yield indications of the advantages of word prediction. A more strictly performed single-case study with an alternating experimental design and randomized allocation of writing conditions, as described in Todman & Dugard, (1999, 2001), would produce a more statistically valid result. The design should allow for differing characteristics of the individual writer, such as original speed, writing strategy regarding aspects such as planning, number of extra key presses and revisions, and linguistic skills in areas such as text construction, spelling and decoding.

Final remarks

What can be said about a word prediction program in general is that it is meant to reduce the number of keystrokes, enhance writing speed, and give support for linguistic functions, such as spelling and sentence construction. The purpose of this study was to investigate the technical effectiveness of prediction techniques and Prophet II in particular. Even though the goal of keystroke savings was immediately achieved, this writer with dyslexia took 13 months to reach his original writing speed, and only when the word prediction was combined with a speech synthesis module. However, his main goal was to produce qualitatively better texts by means of linguistic writing support. [2] The argument *against* the word prediction solution has been that it slows down the writing speed for persons with no motor dysfunction. This argument needs to be reconsidered in light of the participant's long-term achievement. For many persons the aspect of typing speed is not of vital importance, but added to the effort (keystroke) savings and the support provided in the process of creating qualitatively better texts, it enhanced the benefit of the prediction method for the person evaluated in this study.

References

Carlberger A & Hunnicutt S (1997). A corpus-based predictive system for writing support in dyslexia. *TMH-QPSR, KTH, Stockholm, 1/1997,* 31 - 35.

Carlberger J (1997). *Design and Implementation of a Probabilistic Word Prediction Program.* Unpublished Master's Thesis in Computer Science, KTH, Stockholm, Sweden.

Ejerhed E, Källgren G, Wennstedt O & Åström M (1992). *The Linguistic Annotation System of the Stockholm-Umeå Corpus Project* (REPORT NO. 33.): Department of General Linguistics, University of Umeå.

Higginbotham DJ (1992). Evaluation of Keystroke Savings across Five Assistive Communication Technologies. *Augmentative and Alternative Communication, 8*(4), 258-272.

Horstmann Koester H & Levine S (1994a). Modeling the speed of text entry with a word prediction interface. *IEEE transactions on rehabilitation engineering, 2*(3), 177-187.

Horstmann Koester H & Levine S (1994b). *Quantitative Indicators of Cognitive Load During Use of a Word Prediction System.* Paper presented at the RESNA '94 Annual Conference, Nashville.

Horstmann Koester H & Levine S (1996). Effect of a Word Prediction Feature on user performance. *Augmentative and alternative communication, 12*(3), 155-168.

Hunnicutt S (1986). Lexical Prediction for a Text-to-Speech System. In: E H L-G Nilsson (Ed.), *Communication and Handicap: Aspects of Psychological Compensation and Technical Aids.*: Elsevier Science Publishers.

Hunnicutt S & Carlberger J (2001). Improving Word Prediction Using Markov Models and Heuristic Methods. *Augmentative and Alternative Communication.*

Laine CJ & Follansbee R (1994). Using Word-Prediction Technology to Improve the Writing of Low-Functioning Hearing-Impaired Students. *Child Language Teaching and Therapy, 10*(3), 283 - 298.

Magnuson T (1994). Evaluation of "Predict": An investigation and a follow-up study of a Swedish word prediction program. *STL-QPSR, TMH, KTH, Stockholm* (4), 1-20.

Magnuson T & Carlberger A (1997). *Profet II, a New Generation of Word Prediction: An Evaluation Study*. Paper presented at the AAATE, A conference on Advancement of Assistive Technology, Thessaloniki, Greece.

Magnuson T & Hunnicutt S (2000). Linguistic Evaluation of Profet II: A Pilot Project. *Logopedics, Phoniatrics, Vocology, 25*(3), 105-114.

Newell A, Arnott J, Booth L, Beattie W, Brophy B & Ricketts I (1992). Effect of the "PAL" Word Prediction System on the Quality and Quantity of Text Generation. *Augmentative and Alternative Communication, 8*(4), 304 - 311.

Newell A, Arnott J & Waller A (1992). On the validity of user-modeling in AAC: Comments on Horstmann & Levine (1990). Letters to the Editor. *Augmentative and Alternative Communication, 8*(2), 89-92.

Palazuelos Cagigas S, Godino Llorente J & Aguilera Navarro S (1997). *Comparison between adaptive and non-adaptive word prediction methods in a word processor for motorically handicapped non vocal users.* Paper presented at the AAATE, A conference on Advancement of Assistive Technology, Thessaloniki, Greece.

Palazuelos-Cagigas SE, Aguilera-Navarro S, Rodrigo-Mateos JL, Godino-Llorente JI & Martín-Sánchez JL (1999). Considerations on the Automatic Evaluation of Word

Prediction Systems. In JC Filip, T Loncke, Helen H Arvidson & Lyle L Lloyd (Ed.). *Augmentative and Alternative Communication: New Directions in Research and Practice*. London: Whurr Publishers.

Soede M & Foulds RA (1986). *Dilemma of prediction in communication aids.* Paper presented at the The 9th RESNA conference, Washington D.C.

Todman J & Dugard P (1999). Accessible randomization tests for single-case and small-n experimental designs in AAC research. *Augmentative and alternative communication, 15*(1), 69-82.

Todman J & Dugard P (2001). *Single-case and small-n experimental designs. A practical guide to randomization tests.* London: Lawrence Erlbaum associates.

Tyvand S, Endestad T, Pedersen D & Heim J (1994). *Improved User Interface for Word Prediction.* Paper presented at the RESNA '94 Annual Conference, Nashville, Tennessee.

Waller A, Beattie W & Newell AF (1991). The computer is mightier than the sword. *Speech Therapy in Practice*, 18-20.

Venkatagiri HS (1993). Efficiency of Lexical Prediction as a Communication Acceleration Technique. *Augmentative and Alternative Communication, 9*(3), 161-167.

[2] Evidence that he achieved his goal of producing qualitatively better texts was given in 9. Magnuson T, Hunnicutt S: Linguistic Evaluation of Profet II: A Pilot Project. *Logopedics, Phoniatrics, Vocology* 2000;25 (3): **105-114**..