# THE EFFECTIVENESS OF WORD PREDICTION

Heidi M. Horstmann and Simon P. Levine
Rehabilitation Engineering Program, Department of Physical Medicine and Rehabilitation
University of Michigan

## Abstract
Comparisons of user speed on AAC systems with and without word prediction support the view that the cognitive and perceptual processes involved in use of word prediction may counteract the gains in selection efficiency. These processes and their associated time requirements are analyzed, using results from the field of human-computer interaction, as a first step toward a comprehensive understanding of word prediction's effectiveness. The ultimate goal of this work is to develop an analytical modeling framework that can be applied to any user-system interface.

## Background
The past decade has witnessed a tremendous amount of clinical and research interest in the use of word prediction as a means of enhancing rate in augmentative communication (AAC) systems. While there are a number of different types of word prediction, the basic technique takes advantage of the redundancy in the English language to predict a set of words that are the most likely candidates for user entry. Word prediction choices are typically displayed in a short list and are refined as the user inputs additional letters. If the desired word is found in the list, it can be selected with one additional input, eliminating the need to select each letter individually.

Since the development of the first prototypes, much effort has gone into improving the predictive algorithms (1,16) and commercializing the systems for widespread use (5). Current systems employ improved algorithms which are capable of learning the pattern of a user's word usage. Future systems promise to go beyond statistical usage tables to utilize inferences about semantics and pragmatics (18).

## Statement of Problem
A critical question is: Just how much performance improvement can be expected using word prediction? Unfortunately, the answer is that nobody knows for sure. Many researchers have tried to address the question by comparing the efficiency of systems with and without word prediction. Common metrics of comparison are keystroke savings for direct select interfaces (16), or combined savings in scan steps and switch hits for scanning interfaces (3). Theoretical simulations show efficiency gains of 24-70% (3,16), depending on the characteristics of the entered text and the word dictionary. Maximum efficiency has been estimated at 82%, under the assumption that each word can be selected with only one keystroke (16). Studies on actual users report efficiency gains of 23-47% (1,13). Based on these data, it's tempting to conclude that word prediction will universally enhance a user's text entry rate by at least 25% and possibly 50%.

However, comparisons of text entry rate with and without word prediction show that word prediction techniques do not work as well in practice as they do in theory. Experimental measurements on subjects with disabilities have demonstrated no improvement (13) or only modest (3-10%) improvements (3). At least one clinical case study has confirmed that while efficiency may improve significantly, text entry speed may not (17).

These data quantitatively confirm what has long been known: decreasing the number of necessary selections may increase the time required to make each selection, leading to unknown effects on overall performance (4,7,14). This does not mean that word prediction *never* enhances rate. It merely points out that the claims that word prediction is "time saving" (5), "increases typing rate" (6), and provides "quicker access" (12) need to be examined more closely in order to determine when they hold true and when they do not.

## Approach
There are three main questions that need to be answered in order to rigorously understand the trade-off between efficiency and selection time. These are:

1. What are the factors that contribute to an increase in selection time?
2. What are the time requirements associated with each of these factors?
3. Can we determine a cross-over point between rate enhancement and rate inhibition by integrating this information with the efficiency data?

While the definitive answers to these questions require further research, there is a significant body of literature in the field of human-computer interaction (HCI) that can be fruitfully applied today, as demonstrated below.

### *Factors Affecting Selection Time*
Use of a word prediction feature requires additional cognitive and perceptual processes, and these are the major contributors to the increase in selection time. Processes that are frequently cited include the visual search of the prediction list and the subsequent decision about whether the list contains the desired word (7,14,17). Additionally, an often overlooked source of cognitive load is the processing involved in planning use strategies and guiding overall activity (2,9). For example, the user may spend time deciding whether or not to search the list at all. Not all users will employ this strategy, choosing instead to either search every time or not at all; however, many users

may adjust their use of word prediction based on their perception of the recent success rate (17). As a second example, some users may exhibit noticeable delays if the word is not found in the list, which may correspond to the processing required to shift the current task from one of recognition to text generation (17).

Word prediction may also affect the act of physically making a selection, once all decisions have been made. In the case of a keyboard-based system, motoric time may increase if the keys added for selection of word prediction choices are significantly more difficult for the user to access. In a scanning system, motoric time depends on the switch hit time as well as the scanning distance to the selection, and this distance is likely to be reduced using word prediction.

*Associated Time Requirements*

The times required for each of these component actions may vary widely between AAC users. However, many of these processes can be quantified based on the experimental performance of able-bodied subjects; this approach provides a "best-case" baseline and can be expected to apply to AAC system users who have cognitive and perceptual abilities within normal limits.

Several HCI studies have been performed on the visual search of lists, in which subjects search for a given target word and make some motor response to choose the target (8,15). The response times reported, when corrected for the motor time, provide estimates of the time spent in visual search and target recognition. For short lists (around five items), ordered either alphabetically or by frequency of use, search times after some practice are 1.0-1.5 seconds (8,15) and may be expected to increase logarithmically if more items are added to the list (8). With substantial practice, it may be possible to achieve search times of 0.5 seconds, although this estimate has not been validated (7).

The other cognitive processes discussed above are not as directly quantifiable, although relevant HCI work can supply approximations here as well. First, since the user of a word prediction system is faced with a choice of text generation methods (either to search the list or ignore it), times measured for choosing between methods in other domains indicate how long this decision might take. In a study of expert spreadsheet users, subjects consistently took an average of 1.76 seconds just to choose whether to type or point to cells in entering a formula (9). Other work has estimated cognitive processing of this type to take 0.62-1.35 seconds (2). Second, the amount of time required to shift attention from word recognition to text generation can be estimated using an established model of human information processing (2). While the details of the model are beyond the scope of this paper, the cognitive shifting task can be considered to require one cycle of the "cognitive processor" (2), or about 0.1 seconds, for a skilled user without cognitive deficits. While these specific estimates are little more than educated guesses at this point, the basic concept -- that unobservable cognitive processes take measurable and sometimes lengthy amounts of time -- has been well-validated in studies of HCI (2,9,10).

*Finding the Cross-over Point*

A primary goal is to establish methods that can define the cross-over point between rate enhancement and rate inhibition in terms of system parameters and user characteristics. One approach to this is to gather more data on users' speed with and without word prediction and attempt to draw some general conclusions based on the results. However, while this approach may successfully determine cross-over points for specific user and system characteristics, it cannot make predictions about how changes in either the user or the system will affect the cross-over point.

A more comprehensive approach attempts to create an analytical framework that integrates system and user factors and supports the simulation of unlimited user-system combinations. Preliminary work has demonstrated the use of one such framework to make theoretical predictions about user performance with and without word prediction (7). Model simulations predicted that text entry speed with word prediction would usually be lower than speed using letters only, using parameter values like those discussed above. Other investigators have also explored analytical modeling techniques, with similarly interesting results, that demonstrate the potential power of the modeling approach to address the numerous trade-off issues that exist in AAC (e.g., 3,4).

The following example illustrates one type of simple analysis that can be done with modeling, using the timing parameters estimated above. First, for a keyboard-based letters-only system, text entry speed can be estimated at $(5.7)(T_k)$ seconds/word, given 5.7 letters/word and $T_k$ as the user's keypress time. If word prediction is added, with a keystroke savings of 50%, text entry speed becomes $(2.85)(T_{cp} + T_k)$, where $T_{cp}$ is the time spent on cognition and perception for each selection. Assuming $T_k$ is the same for both systems, the equations predict that the letters-only system will be faster for all $T_k < T_{cp}$. Using best-case timing values, and assuming all processes occur in series (2), $T_{cp}$ is 1.22 seconds, which implies that word prediction will not enhance rate for individuals whose keypress time is less than 1.22 seconds. Note that this example is only an illustration of the approach, and continued research is necessary before specific values can be applied in practice.

**Implications**

A major implication is the need for a shift in research focus from system development to user-system interaction (11). Specific questions that need attention include:

1. Do users actually perform all the processes discussed above? If not, why not?
2. How much time does each process really take? And what is the individual variation?

3. How do the times add up? Are processes performed serially, or partially in parallel?
4. How well do model simulations predict actual performance?

Answers to these questions require a great deal of empirical measurement and observation. The key point is that, in addition to measuring the overall performance, we assess the individual contributions of the component actions that produce that performance, in order to build a foundation for a general modeling technique.

Although much research remains to be done, these ideas have practical implications today. For AAC clinicians and users, an awareness of the cognitive and perceptual costs that may be introduced with word prediction provides an important balance to manufacturers' claims and can help the user make a more informed decision. For system developers, application of the ideas within the relevant HCI literature could result in significant design improvements.

**Discussion**

An important limitation of this approach is that while its primary focus is text entry rate, there are numerous additional factors that determine the ultimate success of any AAC system. For example, users may express preference for a word prediction system because it helps their spelling, regardless of its effects on sheer speed (17). Additionally, improving physical efficiency may reduce fatigue for some users, allowing them to work longer or more comfortably. Finally, a user may just have a personal preference for a particular system. The optimal mix of text accuracy, user fatigue, and communication speed depends greatly on the specific goals and abilities of the user and achieving this requires a combined effort of clinician and user. However, a framework that provides an understanding of the factors that determine text entry speed and predicts the speed that may be accomplished with practice would provide a significant contribution to this effort.

**References**


1. Bentrup, J.A. (1987). Exploiting word frequencies and their sequential dependencies. *Proc. of 10th Annual RESNA Conf.*, Washington, D.C.: RESNA, 121-123.
2. Card, S., Moran, T., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
3. Gibler, C.D. and Childress, D.S. (1982). Language anticipation with a computer-based scanning communication aid. *Proc. of the IEEE Computer Society Workshop on Computing to Aid the Handicapped*, Charlottesville, VA, 11-15.
4. Goodenough-Trepagnier, C. and Rosen, M.J. (1988). Predictive assessment for communication aid prescription: motor-determined maximum communication rate. In Bernstein, L. (Ed.), *The Vocally Impaired*. Philadelphia: Grune and Stratton.
5. Gorgens, R.A., Bergler, P.M., and Gorgens, D.C. (1990). HandiWARE. *Proc. of RESNA 13th Conf.*, Washington, D.C.: RESNA, 43-44.
6. Gunderson, J. and Vanderheiden, G. (1988). One-screen multiplexed keyboard for transparent access to standard IBM PC software. *Proc. of ICAART 88*, Washington, D.C.: RESNA, 378-379.
7. Horstmann, H.M. and Levine, S.P. (1990). Modeling of user performance with computer access and augmentative communication systems for handicapped people. *Augmentative and Alternative Communication, 6:4*, 231-241.
8. Landauer, T.K. and Nachbar, D.W. (1985). Selection from alphabetic and numeric menu trees using a touch screen: breadth, depth, and width. *Proceedings of CHI '85*, New York: ACM, 73-78.
9. Olson, J.R. and Nilsen, E. (1988). Analysis of the cognition involved in spreadsheet software interaction. *Human-Computer Interaction, 3:4*, 309-350.
10. Olson, J.R. and Olson, G.M. (1990). The growth of cognitive modeling in human-computer interaction since GOMS. *Human-Computer Interaction, 5*, 221-265.
11. Rosen, M.J. (1989). Editorial. *Assistive Technology*, 1:79-80.
12. Rowley, B.A. (1988). Interfacing the disabled to computer software through virtual interfaces. *Proc. of ICAART 88*, Washington, D.C.: RESNA, 380-381.
13. Scull, J. and Hill, L. (1988). A computerized communication message preparation program that "learns" the user's vocabulary. *Augmentative and Alternative Communication, 4:1*, 40-44.
14. Soede, M. and Foulds, R.A. (1986). Dilemma of prediction in communication aids and mental load. *Proc. of 9th Annual RESNA Conf.*, Washington, D.C.: RESNA, 357-359.
15. Somberg, B.L. (1987). A comparison of rule-based and positionally constant arrangements of computer menu items. *Proc. of CHI+GI '87*, New York: ACM, 255-260.
16. Swiffin, A.L., Arnott, J.L., Pickering, J.A., and Newell, A.F. (1987). Adaptive and predictive techniques in a communication prosthesis. *Augmentative and Alternative Communication, 3:*, 181-191.
17. Treviranus, J. and Norris, L. (1987). Predictive programs: writing tools for severely physically disabled students. *Proc. of 10th Annual RESNA Conf.*, Washington, D.C.: RESNA, 130-132.
18. Yang, G., McCoy, K., and Demasco, P. (1990). Word prediction using a systemic adjoining grammar. *Proc. of 13th Annual RESNA Conf.*, Washington, D.C.: RESNA, 185-186.


**Address**


Heidi M. Horstmann
Rehabilitation Engineering Program
1C335 University of Michigan Hospital
Ann Arbor, Michigan 48109-0032, USA