Proceedings of the Fourth IASTED International Conference
Telehealth and Assistive Technologies (Telehealth/AT 2008)
April 16-18, 2008 Baltimore, Maryland, USA

WORD PREDICTION AND COMMUNICATION RATE IN AAC

Keith Trnka, John McCaw, Debra Yarrington, Kathleen F. McCoy
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716
email: {trnka,mccaw,yarringt,mccoy}@cis.udel.edu

Christopher Pennington
AgoraNet, Inc.
314 East Main Street, Suite 1
Newark, DE 19711
email: penningt@agora-net.com

ABSTRACT

Word prediction systems can reduce the number of keystrokes required to form a message in a letter-based AAC system. It has been questioned, however, whether such savings translate into an enhanced communication rate due to the additional overhead (e.g., shifting of focus and repeated scanning of a prediction list) required in using such a system. Our hypothesis is that word prediction has high potential for enhancing AAC communication rate, but the amount is dependent in a complex way on the accuracy of the predictions. Due to significant user interface variations in AAC systems and the potential bias of prior word prediction experience on existing devices, this hypothesis is difficult to verify. We present a study of two different word prediction methods compared against letter-by-letter entry at simulated AAC communication rates. We find that word prediction systems can in fact speed communication rate (an advanced system gave a 58.6% improvement), and that a more accurate word prediction system can raise communication rate higher than is explained by the additional accuracy of the system alone due to better utilization (93.6% utilization for advanced vs. 78.2% for basic).

KEY WORDS

Natural Language Processing, Word Prediction, AAC, Communication Rate

1 Introduction

In the United States alone, it is estimated that approximately two million people are afflicted with a speech disability severe enough to create a difficulty in being understood [3]. This disability affects these individuals in all aspects of their lives, from education, to the workplace and into their personal lives. In an effort to aid these individuals in their basic need of expressing their thoughts, high-tech AAC (Augmentative / Alternative Communication) has been developed. This is a multi-disciplinary field combining the ideas from Electrical Engineering, Computer Science, Linguistics, Disability Studies, Speech Pathology and Cognitive Sciences to create specialized computers that assist people in communicating.

In this work we concentrate on people who have no cognitive or linguistic impairments, but use AAC because of motor impairments which render them unable to speak in an understandable fashion. Motor impairments can make entering letters and words to be spoken by the AAC device a slow and challenging process. Actual communication rates of AAC users are often below 10 words per minute [11], compared to the common 130-200 words per minute speech rate of speaking people. This creates a large communication gap between AAC and traditional communicators, making for a quality of life issue.

Do you enjoy	c	
	cooking	(F1)
	camping	(F2)
	classical	(F3)
	comedy	(F4)
	college	(F5)

Figure 1: Sample word prediction interface

Because of this, communication rate is one of the most acknowledged problems in AAC interactions [12, 3]. In this work, we focus on devices that use a physical interface of keys for letters, and attempt to minimize the number of keystrokes needed to input a word by using word prediction.

Word prediction is an interface technique that relies on the user's inputted characters to guess the word the user is attempting to generate. If the system were to correctly guess the intended word, the user could select it at the cost of one keystroke instead of continuing to type out the word letter by letter. An example word prediction system is shown in Figure 1. The words presented in the list all begin with the characters that the user has already entered. Intuitively, word prediction systems can save significant time entering words and therefore it has been implemented in many of today's common AAC devices including Prentke Romich's Pathfinder, Dynavox Technology's Dynavox 4 and Saltillo's ChatPC.

Despite the intuitive appeal of word prediction, several researchers have found that word prediction gave only minor improvements in communication rate at best and lowered communication rate at worst [15, 7, 1]. They hypothesize that the cognitive and perceptual load of word prediction (e.g., scanning the list, recognizing the word) consume all of the time saved by using fewer keystrokes.

However, previous studies made certain design choices that may influence the results. Venkatagiri [15] collected 10 natural sentences from 16 participants, but the word prediction entry of the sentences was performed by a single user with no motor impairments. The user was expected to repeatedly scan 15 predictions and select the word as soon as it appeared. They found that the difference in communication rate with word prediction was not significantly different than letter-by-letter entry, despite nearly 50% keystroke savings. A possible explanation for this finding could be due to the combined overhead of a very large prediction list with having the user scan the predictions after every key.

Koester and Levine [7] performed a similar experiment, but additionally studied the differences between two different word prediction user strategies. In the first strategy, participants were instructed to scan the predictions after every key press. In the second strategy, the participants were instructed to not look at predictions for the first two characters, and then begin to scan the list after every key press. Participants were selected from people with spinal cord injuries (SCI) and also a sample of people without motor impairments, using either a head stick or a hand splint to access the keyboard. They found that the group without motor impairments displayed a moderate increase in communication rate, while the participants with a spinal cord injury displayed a drastic decrease in communication rate. One of the reasons for the large drop is prior experience. The SCI group were experts at letter-by-letter entry compared to the other group, but slowed down to the rate of the novices when presented with word prediction.

Anson et al. [1] studied usage of an on-screen keyboard using a mouse. The 10 participants had no motor impairments, but the communication rates achieved with and without word prediction are comparable to users of AAC devices. On average, participants' communication rates improved by roughly 10% through use of word prediction over letter-by-letter entry, but many participants were frustrated with the predictions. A dictionary-based prediction system may have been responsible for the mediocre improvement and user frustration (in fact, their results are comparable to our basic prediction method).

Venkatagiri [15] and Koester and Levine [7] instructed participants on when to scan the predictions and how to use the system. In contrast, users of text entry systems naturally develop their own strategies for optimizing communication rate. People who type very quickly will likely find that word prediction is a distraction. People who type slowly may find that the keystroke savings of word prediction outweighs the cognitive and perceptual overhead of using the system. However, even slow typists are unlikely to fully utilize the system (i.e., select the word as soon as it appears), rather, we expect that users will sometimes find it faster to type a word out instead of scanning the list at every opportunity. In contrast to previous studies, we allow users to develop their own strategy of usage to optimize communication and allow them to either use or ignore word prediction information as they wish.

Under the belief that word prediction can benefit users, many researchers are currently investigating ways to improve word prediction to save more keystrokes [9, 10, 14, 2, 6]. These methods include predictions based not just on characters entered, but also on previous whole words entered and even topics that a user is talking about. With these advances in prediction modeling, the question arises as to whether or not such advances would help word prediction become more effective. In our previous study with 17 participants [13], we found that an advanced prediction method significantly increased communication rate well beyond that which can be explained by the method's increase in accuracy alone.

This paper is a continuation of our previous study using 28 participants, which lends itself to a stronger statistical analysis and more interesting correlations. In the study we compare three different *text entry methods* (also referred to as *input methods*) for AAC text: typing on a keyboard with no word prediction, typing on a keyboard with a basic word prediction algorithm available, and typing on a keyboard with a more advanced word prediction algorithm available. In these three input situations we investigate (1) the effect of the entry method on input rate (i.e., the time it takes to enter a single keystroke) and (2) the effect of the entry method on communication rate. We will also discuss (3) the prediction utilization of the algorithms and (4) survey results from participants under each condition.

We hypothesize that despite a decrease in input rate due to increased cognitive load, a more advanced word prediction algorithm will allow the user of an AAC system to communicate faster. We also expect that a more advanced word prediction algorithm will increase the user's trust in the system, enticing the user to use it more frequently.

Although the target population of our research is adult AAC users without significant cognitive impairments, undertaking a significant study using the target population would be infeasible and potentially biased. Devices are often customized to suit each particular AAC user, both in the physical interface as well as the device software. Developing a word prediction system that interoperates with each user's preferred typing environment is impractical. But more importantly, many AAC users are already familiar with some form of word prediction system in commercial devices. If the predictions are poor in the user's system, they will likely be biased to not utilize another word prediction system. Therefore, we chose to study adults who had little or no prior experience with word prediction and no motor or communication difficulties, but were slowed using a key press delay to simulate average AAC communication rates. This helps us to eliminate the bias of previous word prediction experience and to allow a consistent user interface for all participants. The results of this experiment have provided us with evidence that motivates continued work on improving word prediction systems for AAC. We plan to validate these results with a small group of AAC users in the future.

Figure 2: View of the copy task with word prediction

2 Experiment

The purpose of this study was to measure and compare the communication rate generated from varying AAC-like systems, some using word prediction and others not. To accomplish this, we compared 3 different systems in the task of copying text into a computer: a somewhat typical word prediction system ("basic"), word prediction using a more language-based algorithm ("advanced"), and no word prediction.

2.1 Text Entry

Each participant was asked to copy text into the computer using the on-screen Wivik™ keyboard and a touch screen, shown in Figure 2. We selected a copy task to control for message composition time, and avoided auditory prompts due to potential complications with spelling.

To simulate AAC users, the Wivik™ Keyboard would not accept another key for 1.5 seconds after a key was pressed. This same delay was implemented in our prediction list as well. With the assumption that words are 5 characters long, this creates an ideal typing speed of approximately 8 wpm, which is consistent with available statistics for AAC [11]. The prediction list is not updated until after 1.5 seconds, so the user could not artificially lower their cognitive overhead.

Our system consisted of 4 active windows with a black desktop background to minimize user distractions. The text to copy was presented at the top of the screen. A new sentence would appear once the user hit the large red "enter" key in the lower right corner of the keyboard. The middle window showed what the user was currently typing. The on-screen keyboard at the bottom of the display was used for letter-by-letter typing. In the cases where word prediction was used, a list of predicted words appeared on the bottom left. This list consisted of the top 5 words generated by the algorithms detailed below. A word from this list could be selected by touching it and waiting 1.5 seconds for the delay. In the case when no word prediction was used, no such box appears.

2.2 Word Prediction Methods

The **basic prediction algorithm** orders the prediction list based on how often that word was used in that session up until that point. If the window's 5 word spots are not filled from this recency model, then the remaining spots are filled from a large dictionary in alphabetical order. While many AAC systems use a more sophisticated algorithm (taking a word's frequency into account as well), the recency algorithm we employed is simple to implement and provides a reasonable approximation of a basic algorithm employed by AAC word prediction systems. We also hope that, if (as hypothesized) the quality of predictions affected communication rate, the difference in the two methods would show a difference in communication rate.

The **advanced prediction algorithm** employs statistical natural language processing techniques to achieve more accurate predictions. The primary prediction method is a trigram language model with backoff, a common statistical technique in Computational Linguistics [8, 4, 16]. When using such a model, the probability of a word appearing in the prediction list is dependent on the preceding two words. This method relies on statistics generated from a training corpus — in our case the Switchboard Corpus, which consists of approximately 2.6 million words.[1] Thus the probability of a particular word being predicted depends on the number of times in the training corpus that the word followed the two words the user has just typed.

For example, suppose the phrase "for a walk" is seen often in the training phase. If the user inputs the text "for a", the word "walk" should appear near the top of the prediction list. However, many three word combinations occur infrequently, so the system also considers the previous word entered by the user as well. For example, if the system frequently encounters the phrase "peanut butter", and the user has entered "peanut", then "butter" should appear near the top of the word prediction list. The final step in this ngram backoff model is to also account for word counts, so that the commonness of a word positively influences its ordering if the context is unknown. These three models, known as trigram, bigram and unigram models are combined to populate the prediction window. If the prediction window does not consist of 5 words at this time, then the recency model and dictionary from the basic prediction algorithm are used.

2.3 Participants

The study was conducted using 28 adult participants with no visual or cognitive impairments and who are native speakers of American English. Most of the participants were obtained through the use of flyers and an auction-like site for college students. Participants were compensated $30 for their time and effort.

[1] The conversations used for the copy task are not included in the training of the model.

2.4 Methodology

Each participant completed 3 sessions, each lasting less than one hour. Each session was separated by at least one day, but no more than 1 week. In each session, participants were asked to type one of three transcribed conversations from the Switchboard Corpus into our computer system using one of three text entry methods (no prediction, basic prediction, advanced prediction). All participants received the same conversation order, that is, each participant entered Conversation A the first day, B the second, and C the third. Each participant was asked to use a different entry method for each day. The order of text entry methods was balanced to control for learning effects and bias.

Before each session, the participants were given a sample text to type to familiarize themselves with the on-screen keyboard and text entry method. Afterwards, they began entering the conversation. Each conversation was given to the participant one sentence at a time to reduce confusion. The experimental software time and date stamped relevant information about each keystroke in a log file. Information about the words appearing in the word prediction list was also logged. After completing each test, the participant was given a short survey to capture their opinions about the text entry method.

The Switchboard Corpus was chosen for this experiment because it was thought to be the closest text available to AAC text. This is a collection of phone conversations between two speakers that was transcribed by humans. In order for the text to be more natural, we removed any speech repairs. The participants were instructed to copy the given text and were not pressured to use the word prediction system, but were encouraged to copy the text quickly and accurately.

3 Results

Once the participants completed the experiment, we processed the resulting log file to assess performance. We analyzed the data by averaging across participants within each of the three different text entry methods. The variables of interest for each method include the input rate (time to press keys), communication rate (words produced per minute), task completion time, and interaction with predictions (keystroke savings, prediction utilization). Each of these statistics were computed for each session, and then averaged over all logs that used the same text entry method.

3.1 Input Rate (seconds per keystroke)

The input rate is measured using seconds per keystroke (spk), which measures the amount of time that it took a user to hit one key. As previously shown [13, 15, 7], the input rates of users when a word prediction system is available is significantly slower than when it is not. What is interesting to note is our experiment is that the time to select a key using advanced word prediction was also significantly higher than when using basic predictions. As shown in Figure 3(a), the input rate for users using the advanced algorithm is 2.95 spk, compared to the 2.59 spk of the basic algorithm and 2.26 spk when no word prediction is given. The differences between the methods' input rates are all highly significant ($p < .0005$).

From a closer look at the data, we found that users took more time to select a predicted word than select a letter from the keyboard. We also found that the time to select a letter from the keyboard was higher with basic prediction vs. no prediction and higher still with advanced prediction.

3.2 Communication Rate (output rate)

The statistic of most interest to AAC users and speech language pathologists alike is communication rate. The communication rate is measured by the number of words a user produces in a minute (wpm). As hypothesized, there is a significant difference between the communication rate of users when using an advanced prediction algorithm over the basic prediction model and no prediction ($p < .0005, .0005$). Additionally, communication rate using basic prediction is significantly faster than no prediction ($p < .0005$). As illustrated in Figure 3(b), users communicate on average 45.4% faster using advanced prediction over the basic one and 58.6% faster than using no word prediction.

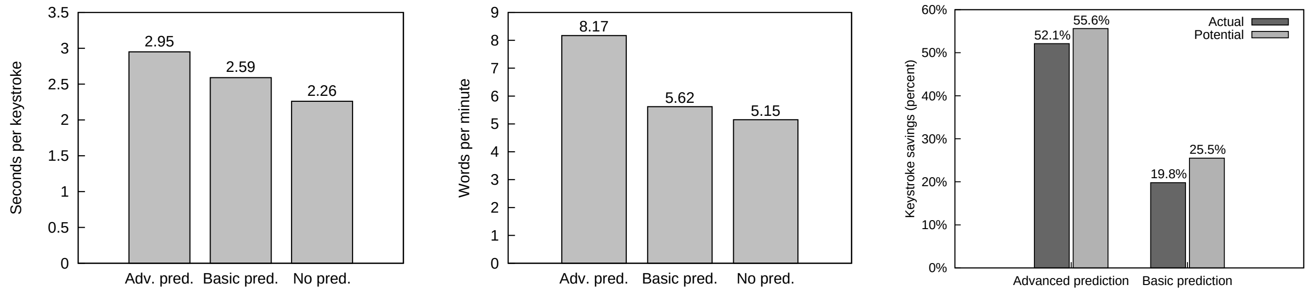
The average total copy task time using the advanced algorithm was 21 minutes, 36 seconds, which was much quicker than it took the participants to complete the task with the basic algorithm (29 minutes 47 seconds) and no prediction (32 minutes 59 seconds).

3.3 Keystroke Savings (potential vs. actual)

Potential keystroke savings is the keystroke savings the user would have achieved if they had used the prediction system fully. On the other hand, *actual keystroke savings* is the amount of characters that the user actually saved by using word prediction. This is normally lower than potential keystroke savings because the user may choose to not check the predictions after every letter and periodically miss the desired word. When using advanced word prediction, as shown in Figure 3(c), actual keystroke savings was 52.1%, compared to the potential savings of 55.6%. Using the basic algorithm, users had an actual keystroke savings of 19.8% and a potential savings of 25.5%.

3.4 Prediction Utilization

Prediction utilization is the actual keystroke savings divided by the potential keystroke savings. This represents how much participants depended on the system for text entry. A user that trusts the system's predictions is likely to inspect the list often, whereas a user that finds the system unreliable or distracting will tend to only scan the prediction list when they feel it may offer a great benefit, such as particularly long words or when they expect a word to



(a) Input rate (b) Communication rate (c) Keystroke savings (actual vs. potential)

Figure 3: Results, from left to right: input rate in seconds per keystroke, communication rate in words per minute, and a comparison between the actual and potential keystroke savings

be predicted. When using the more advanced prediction algorithm, users had a prediction utilization of 93.6% versus 78.2% utilization for the more basic method. These percentages are statistically different ($p < .0005$).

3.5 Survey Results

The exit survey consisted of 12 questions using a 5 point Likert-type scale and 5 open ended questions. We report the findings from just a few of the questions which are relevant to this paper. All of the reported results were statistically significant ($p < .05$).

Fatigue is particularly relevant to actual AAC users, as a full day of communicating using an AAC device is often tiring, which in turn tends to reduce communication rate. When asked how tiring the method was, users felt that advanced word prediction was much less tiring than basic word prediction and no prediction. Therefore improvements in word prediction are likely to reduce AAC user fatigue, which in turn should help AAC users maintain a steady communication rate throughout the day.

When asked about ease of use, users felt that the more advanced prediction system was much easier to use than the basic prediction system. Similarly, when asked about the speed of entry, users perceived the increase in communication rate due to better predictions. Also, users found advanced word prediction more useful than basic word prediction. In response to the worst thing about basic prediction, one user said, "none of the words seemed to match what I wanted to say."

Question 11 asked a somewhat different question than the others: "How distracting did you find the word prediction?" Seemingly counter-intuitively, users said that they found advanced word prediction less distracting than basic prediction, even despite the significantly higher cognitive overhead of advanced word prediction (shown in Section 3.1). If the input rate was slower, spending more time making decisions based on the word prediction list of the advanced system is to blame. This decision time should create a distraction for the users. However, our users equated distraction as any effort that did not help accomplish their task, essentially combining cost (cognitive overhead) and benefit (keystroke savings).

4 Discussion

The major finding of this study and our previous work [13] is that a good word prediction system can improve communication rate, more so than a basic system. The increase in communication rate is not only due to increased potential keystroke savings, but users also utilize the system more fully when they trust the predictions.

We found that input rate decreased when any word prediction was available, and more so with advanced prediction. This increase in cognitive load is consistent with our previous findings as well as other researchers [13, 15, 7]. The decrease in input rate due to word prediction is explained by an increasing reliance on the prediction system. If a user trusts the system more, he or she will scan the prediction list more often, increasing cognitive load. However, this increased cognitive load was *chosen* by our users — they found that the cost of added cognitive load was outweighed by the benefit of the keystroke savings offered by word prediction, especially for advanced prediction. The exit survey results also confirm that users accept the added cognitive load in return for a faster communication rate. Most users found that advanced predictions were very useful, and were significantly more useful than basic predictions. Question 11 again shows users saying that the benefit of word prediction is worth the added overhead.

The real-world benefit of word prediction is twofold: first, word prediction significantly increased communication rate. Secondly, word prediction reduced user fatigue. Advanced predictions increased communication rate by 58.6% over no word prediction and 45.4% over basic word prediction. This increase in communication rate occurred despite a significant decrease in input rate — the added keystroke savings of word prediction far outweighed the cost of a slower input rate. Not only did log analysis show this, but also the users could notice the difference in the exit survey. Additionally, the survey found that users felt significantly less fatigue when using advanced word prediction as compared to basic prediction and no prediction. Users found that even basic prediction reduced their fatigue. These trends indicate that additional improvements in the potential keystroke savings of word prediction are likely to further increase communication rate and decrease user fatigue.

The more semantically and syntactically correct predictions increased user trust in our system, as hypothesized by [5]. When asked to describe what he or she liked best about advanced prediction, one participant noted:

> I liked how depending on the context of the sentence ... the word prediction would come up with relevant words that would often be correct for the next word I needed to type.

The increased benefit of more accurate predictions is noted not only in the surveys, but also in the log analysis — advanced prediction offered 55.6% keystroke savings and users realized 93.6% of the system's potential — 52.1% keystroke savings. However, the basic prediction offered only 25.5% keystroke savings, and users trusted the system less, scanning the predictions less often — users only utilized 78.2% of the potential savings of basic prediction. In addition to the significant effect of user trust on prediction utilization and in turn, actual keystroke savings, we found that there were some variations in the utilization of different users, where some users seemed to not like word prediction or avoid it due to previous experiences.

5 Conclusion

In this paper we have extended our previous study [13] to show that word prediction, when implemented properly, can lead to higher communication rates and less fatigue than poorly developed word prediction systems and letter-by-letter entry. We have empirically verified the increased cognitive load due to word prediction, but have also shown that the increased keystroke savings compensates for the added overhead and that participants more fully utilize a better word prediction system. This paper has shown that further research and development in word prediction can potentially increase the quality-of-life of AAC users. In future studies we hope to further extend this research to AAC users and more advanced word prediction methods.

Acknowledgement

This work was supported by US Department of Education grant H113G040051.

References

[1] D. Anson, P. Moist, M. Przywars, H. Wells, H. Saylor, and H. Maxime. The effects of word completion and word prediction on typing rates using on-screen keyboards. *Assistive Technology*, 18, 2004.

[2] C. Beck, G. Seisenbacher, G. Edelmayer, and W. Zagler. First user test results with the predictive typing system FASTY. *ICCHP*, pages 813–819, 2004.

[3] D. Beukelman and P. Mirenda. *Augmentative and Alternative Communication: Supporting Children and Adults with Complex Communication Needs*. Paul H. Brookes Publishing Company, Baltimore, MD, 2005.

[4] P. F. Brown, P. V. deSouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479, 1992.

[5] P. Demasco, K. McCoy, Y. Gong, C. Pennington, and C. Rowe. Towards more intelligent AAC interfaces: The use of natural language processing. In *RESNA*, New Orleans, Louisiana, June 1989.

[6] N. Garay-Vitoria and J. Abascal. Text prediction systems: a survey. *Univ Access Inf Soc*, 4:183–203, 2006.

[7] H. H. Koester and S. P. Levine. Modeling the speed of text entry with a word prediction interface. *IEEE Transactions on Rehabilitation Engineering*, 2(3):177–187, 1994.

[8] I. Langkilde and K. Knight. The practical value of n-grams in generation. In *Proceedings of the International Natural Language Generation Workshop*, Niagra-on-the-Lake, Ontario, 1998.

[9] G. W. Lesher and D. J. Higginbotham. Using web content to enhance augmentative communication. In *Proceedings of CSUN 2005*, 2005.

[10] J. Li and G. Hirst. Semantic knowledge in word completion. In *ASSETS*, pages 121–128, 2005.

[11] A. Newell, S. Langer, and M. Hickey. The rôle of natural language processing in alternative and augmentative communication. *Natural Language Engineering*, 4(1):1–16, 1998.

[12] J. Todman and N. Alm. Talk boards for social conversation. *Communication Matters*, 11:13–15, 1997.

[13] K. Trnka, D. Yarrington, J. McCaw, K. F. McCoy, and C. Pennington. The Effects of Word Prediction on Communication Rate for AAC. In *NAACL-HLT; Companion Volume: Short Papers*, pages 173–176, 2007.

[14] K. Trnka, D. Yarrington, K. McCoy, and C. Pennington. Topic Modeling in Fringe Word Prediction for AAC. In *IUI*, pages 276–278, 2006.

[15] H. S. Venkatagiri. Efficiency of lexical prediction as a communication acceleration technique. *Augmentative and Alternative Communication*, 9:161–167, 1993.

[16] S. Wang, D. Schuurmans, F. Peng, and Y. Zhao. Semantic n-gram language modeling with the latent maximum entropy principle. In *AISTATS*, 2003.