

Effect of Window Size on Rate of Communication in a Lexical Prediction AAC System

Horabail S. Venkatagiri

Department of Speech Communication, Iowa State University, Ames, Iowa, USA

The study evaluated the effect of three different "window" sizes (the number of words displayed in the word selection menu) on the efficiency of communication in a lexical prediction augmentative and alternative communication (AAC) program. The results showed that a 15-word window produces significantly higher predictions and fewer keystrokes than a 5-word window. The time required to type the messages was statistically equivalent across windows. However, mean keystroke duration was significantly higher in the 15-word window as compared to the 5-word window. This study shows that (1) a saving in keystrokes does not necessarily result in an increase in the rate of communication in lexical prediction programs; (2) searching a word prediction window for target words places significant cognitive/perceptual demands on even individuals with presumably normal cognitive and perceptual abilities (and probably much more so on AAC users); and (3) a 15-word window produces roughly the same rate of communication as the 5-word window but at a significantly greater saving in effort (keystrokes/switch activations).

KEY WORDS: acceleration, augmentative and alternative communication (AAC), computer effort, lexical prediction, rate, technology

Two major considerations in the implementation of an augmentative and alternative communication (AAC) system for an individual are (1) communication flexibility and (2) communication rate (Scull & Hill, 1988). Communication flexibility refers to the individual's ability to generate an unlimited number of novel utterances. A unique feature of human linguistic communication is that, barring a few phrases used in routine social interaction, we rarely use an utterance in exactly the same form more than once. Instead, we constantly generate unique utterances — utterances that differ from our own previous speech as well as that of others. Beukelman and Yorkston (1982) have shown that AAC users tend to produce unique sentences in their communication as well.

Rate of communication is also important for the successful use of AAC (Beukelman & Yorkston, 1982; Goodenough-Trepagnier & Rosen, 1988). A slow rate of communication tends to reduce the amount and the overall quality of interaction that takes place between AAC users and their conversational partners.

Unfortunately, flexibility and rate generally work against each other (Damper, 1986). An AAC system that is high on flexibility tends to have a slow rate and vice versa. To overcome this trade-off, several acceleration techniques have been developed (Damper, 1986; Vanderheiden & Kelso, 1987). Word prediction

is one of the most common communication acceleration techniques employed in the microprocessor-based communication programs (Heckathorne, Voda, & Leibowitz, 1987; Horstman & Levine, 1990; Scull & Hill, 1988; Swiffin, Arnott, Pickering, & Newell, 1987).

In its simplest form, the word prediction paradigm requires that the program provide a set of likely words (words weighted for frequency, for example) in response to user's keystrokes. If the word the user wants to type is among the words displayed by the program, the word may be incorporated in the message by typing a code, usually a number, associated with the word. This, of course, results in a saving of keystrokes necessary to complete predicted words.

The disadvantage of the word prediction technique is that the user needs to constantly search computer-generated word lists in order to make selections. This wastes time as well as places additional cognitive demands on the user.

The number of words displayed in the word prediction window probably has an effect on the rate of communication in a lexical prediction communication program. If a small number of words are displayed in the window, many short, less frequently used words may not be predicted, since the words are likely to be completely typed out by the user before they appear. If the window contains a large number of words, the

user is likely to spend a significant amount of time searching the prediction window for the target word, thereby reducing the advantage of word prediction for communication acceleration.

A single window size will probably not work for all AAC users. The cognitive, linguistic, and motor abilities of a user need to be taken into consideration in determining the optimum window size for that individual. However, at present there is little empirical knowledge on which to base clinical decisions about optimum window size in lexical prediction AAC programs. The present study was, therefore, designed to answer the following questions:

1. Do different window sizes yield significantly different rates of communication?
2. Do different window sizes require a significantly different number of keystrokes and hence different amounts of effort?
3. Do different window sizes produce significantly different amounts of word prediction, which affect both rate and effort?

Answers to these questions are of importance to clinicians and AAC users. With the availability of relatively low cost yet powerful notebook-size personal computers and high-quality diphone text-to-speech synthesizers (O'Shaughnessy, Barbeau, Bernardi, & Archambault, 1988), one of the last major obstacles to the implementation of effective AAC appears to be its limited productivity. Productivity here refers to the ability to generate an unlimited number of appropriate responses on a time scale that approximates normal conversation. Researchers are, therefore, turning their attention to questions concerning acceleration techniques in AAC (Damper, 1986).

METHOD

Messenger Program

The program tested in this study was the Messenger (Venkatagiri, 1991), which was written for the Apple II computer with a minimum of 128 kilobytes of random access memory (RAM) and an 80-column capability. It is primarily intended for training in AAC use but can also be used as a stationary communication device. Both direct selection (using the keyboard) and a scan access version are available. The present study used the keyboard access version of the program. Messenger employs the popular, albeit not totally satisfactory (Venkatagiri, 1991), Echo II speech synthesizer for speech output.

Preparation of messages in the program is facilitated by word prediction. The words are predicted on the basis of frequency (of occurrence of words in the language) and recency of usage. At the time of testing this program, the dictionary base for word prediction contained 903 words assembled from various publish-

ed sources of word frequency lists (Yorkston, Dowden, Honsinger, Marriner, & Smith, 1988). The dictionary was created independent of and nearly 1 year before the present study was conceived. The window size (the number of words displayed in the prediction window) is user selectable in the range of 0 (no prediction) to 15.

Stimulus Materials

The sentences used to test the effect of window size were generated by asking a group of 16 college students to write a sentence they might use in each of the 10 hypothetical communication situations: talking to one's friend, mother, father, sister/cousin, brother/cousin, teacher, a counterperson, a stranger, a salesperson, and an acquaintance. Subjects were told that it would be best if they could reproduce utterances they had recently used in one or more of these situations. If that was not possible, they could provide sentences that they were likely to use in these situations. This procedure was used to generate a corpus of realistic utterances, which are likely to be used in commonly encountered communication situations without having to sample a variety of actual conversations.

A total of 160 sentences (16 × 10) were generated in this way. The average sentence length was 6.77 words with a range of 2 to 17 words. Both questions and statements were represented in the corpus, which consisted of 714 content words (nouns, verbs, adverbs, and adjectives) and 370 function words (pronouns, articles, prepositions, and conjunctions). The salient characteristics of the language corpus are presented in Table 1.

Subjects

Twenty-one young adult females, all college students, with a mean age of 22 years, 3 months, participated in the study. All subjects passed an informal speech screening through conversation and a hearing screening at 500, 1000, and 2000 Hz at 20 dB HL. The subjects presumably had normal cognitive, lin-

TABLE 1: Salient Characteristics of the Language Corpus Used in the Study

Number of sentences	160
Mean words per sentence (SD)	6.67 (2.59)
Mean morphemes per sentence (SD)	7.32 (2.89)
Type/token ratio	0.73
Content words	714
Function words	370

guistic, and motor functions. None had a history of neurologic disorders, severe head injury, learning disability, or communication disorder.

Typing skill was integral to the task involved in the experiment. Therefore, the subjects were asked to rate themselves on keyboarding skill on a scale of 1 to 3 (1 = low; 2 = moderate; 3 = high). Five subjects rated themselves as having low keyboarding skill, 10 as having moderate skill, and six as possessing a high level of keyboarding skill. Measures used to control for differences in typing skill are described later in this paper.

The use of only adult, female participants was to ensure homogeneity in subject performance. Gender, age, and other subject characteristics are probably important variables in the successful use of lexical prediction and should be investigated. However, these variables were not the focus of study in the present investigation. Similarly, AAC users were not included in the present study because of the extreme heterogeneity of this population. This was a preliminary study to determine the effect of window size on communication in a group of nondisabled subjects. The findings of this study may form a basis for comparison with clinical populations in future studies.

Procedure

Following an explanation of how messages are typed using the Messenger program, each subject typed a practice sentence to familiarize herself with the experimental procedure and the keyboard layout. The 15-word window size was used during the practice session. The experimental procedure was further clarified during and immediately after this practice session by answering the subject's questions and restating instructions.

Three different window sizes — 5, 10, and 15 words — were utilized in the study. Each subject typed three messages in each of the three window conditions utilizing Messenger's lexical prediction feature for a total of nine trials.

The subjects were instructed to use their preferred hands for typing; all subjects chose to use their right hands. They typed the messages using only their right index finger to ensure that intersubject differences in typing speed were not a confounding factor in the experiment. Additionally, a repeated measures design was used in which all 21 subjects took part in all three experimental conditions to control for typing proficiency differences among subjects.

The messages typed by the subjects were randomly selected from the pool of 160 messages described above (under "Stimulus Materials"). The only restriction placed on random selection was that no message was to be used more than once with any given subject. However, each subject typed a differ-

ent set of sentences to ensure that a variety of messages generated from several different hypothetical communication situations were used in the study.

The order of presentation of the three window conditions was counterbalanced so that an equal number of subjects had each of the conditions as their first, second, and third conditions. This was done to counteract the effect of practice as subjects moved from one window condition to the next.

The subjects received one sentence at a time during typing. Each test sentence, typed on a strip of paper, was placed in front of the subject to type. They were instructed to search the prediction window after typing each letter in the sentence. If they found the target word in the window, they were to select it by pressing the number (0–9) or the special character (such as "[") associated with it. If subjects made a typing error, they were to strike the delete key to erase the error before proceeding.

Dependent Variables

The following dependent variables were computed:

Keystroke Saving Ratio. This ratio is obtained by first dividing the actual keystrokes required to type a message using the word prediction feature by the number of keystrokes that would have been necessary without the prediction feature. The latter value is equal to the number of characters (letters, spaces, and punctuation marks) in the message. The value obtained is then subtracted from one to derive the keystroke saving ratio. A large keystroke saving ratio is obviously desirable.

Duration Ratio. Three duration ratios were computed. In one, the amount of time used to prepare a message (measured in centiseconds, i.e., accurate to one hundredth of a second) was divided by the number of letters in the message (Dur/L). Second, the amount of time needed to prepare a message was divided by the number of words in the message to derive Dur/W ratio. Last, the amount of time used to prepare a message was divided by the number of keystrokes to derive the Dur/K ratio. A small duration ratio results in a higher rate of communication.

A stopwatch was used to measure the total duration of each trial to the nearest hundredth of a second. The stopwatch was started when the subject typed the first letter of the sentence, and the clock was stopped when the subject (or the Messenger program) typed out the last letter of the sentence. For three randomly selected subjects, the experimenter and a research assistant independently measured the time. The mean interexperimenter difference was 9.3 centiseconds with a range of 0 to 19 centiseconds.

Prediction Ratio. This ratio was computed by dividing the number of words predicted in each message by

the total number of words in that message. Thus, if a sentence contained six words and three of these words were completed by the Messenger program, rather than the subject, then the prediction ratio was computed to be 0.5 (50%). A large prediction ratio indicates greater efficiency of the program.¹

Typing Errors

Each typing error made by the subjects added two keystrokes: one that caused the typing error and the second to erase it. There were 38 typing errors (76 keystrokes) out of a total of 3,726 keystrokes executed by the subjects. Window conditions 5, 10, and 15 words had 9, 13, and 16 typing errors, respectively. Typing errors occurred in 27 (14%) out of the 189 total trials (total trials = 3 trials [sentences] × 3 window conditions × 21 subjects). Typing errors occurred in 6 trials of 5-word, 10 trials of 10-word, and 11 trials of 15-word window conditions. Keystrokes associated with typing errors were excluded from the computation of keystroke saving ratios as well as duration ratios.

In computing duration ratios, the time associated with typing errors was excluded in the following manner. First, the mean keystroke duration for the trial in which the error occurred was computed. A duration equal to the mean keystroke duration was then subtracted for each keystroke associated with typing errors on that trial from the total duration of the trial.

Message Length

Table 2 shows the mean lengths and standard deviations of messages measured in words and letters across the three window conditions. The mean difference in length is less than one word and one letter among the three experimental conditions. The standard deviation values are relatively large yet comparable across conditions, showing that each condition contained messages of varied length.

RESULTS

Table 3 shows the group mean and standard deviation values for keystroke saving, duration (Dur/L, Dur/W, and Dur/K), and prediction ratios obtained in the study.

¹Originally, a subject's failure to take advantage of a predicted word was also planned to be used as a dependent variable. However, there were only four instances in which the subjects failed to notice that the word they were typing had appeared in the prediction window: one in the 10-word window and three in the 15-word window conditions. The data was deemed insufficient and dropped from analysis. With AAC users, however, this may prove to be a useful dependent variable.

TABLE 2: Mean Length of Messages across the Three Window Conditions

Window Size	Mean Letters (SD)	Mean Words (SD)
5 Words	32.05 (7.47)	7.33 (2.99)
10 Words	33.50 (9.19)	7.19 (3.16)
15 Words	33.40 (10.30)	6.69 (2.51)

Keystroke Savings

Keystroke savings of about 31%, 38%, and 44% were obtained with window sizes of 5, 10, and 15 words, respectively. A repeated measures analysis of variance (ANOVA) (Winer, 1971) reveals that the means for keystroke savings are significantly different for the three experimental conditions ($F [2,40] = 6.16$; $p < .003$). A post-hoc analysis using the Scheffe multiple comparisons test (SPSS-X, 1988) shows that the difference between the means for window sizes 5 and 15 is significantly different ($p < .05$). The differences in means between window sizes 5 and 10 and 10 and 15, however, are not significant ($p > .05$).

Word Prediction

Mean prediction ratios across window sizes range from 0.588 (58.8%) for the 5-word window to 0.791 (79.1%) for the 15-word window, with an intermediate value of 0.726 for the 10-word window. A repeated measures ANOVA shows that the three window sizes result in significantly different levels of prediction ($F [2,60] = 18.93$; $p < .0001$). A post-hoc comparison using the Scheffe procedure shows that the mean prediction ratios are significantly higher in the 15- and 10-word windows as compared to the 5-word window ($p < .05$). The means for windows 10 and 15 do not differ significantly ($p > .05$).

Letter Duration

The duration ratios for letters (Dur/L) show that the mean time required to type letters using the lexical prediction feature (in which not all letters are actually typed by the subjects) ranges from 1.58 seconds for window sizes of 5 and 10 words to 1.62 seconds for the 15-word window. The values are similar when the duration is computed as a function of words typed. The mean time required to type a word is 6.16, 6.62, and 6.35 seconds, respectively, in window sizes of 5, 10, and 15 words. A repeated measures ANOVA shows that the differences among the three window sizes are not significant for both Dur/L ($F [2,60] = 0.25$; $p > .70$) and Dur/W ($F [2,60] = 0.84$, $p > .40$).

TABLE 3: Means for Keystroke Saving, Duration, and Prediction Ratios

Window	Keystrokes (SD)	Dur/L (SD)	Dur/W (SD)	Dur/K (SD)	Prediction (SD)
5 Words	0.311 (0.163)	158.2 (85.59)	615.79 (130.77)	198.13 (37.18)	0.588 (0.137)
10 Words	0.377 (0.091)	157.5 (94.25)	662.48 (161.36)	217.97 (34.49)	0.726 (0.119)
15 Words	0.439 (0.088)	162.4 (107.07)	634.51 (105.36)	244.00 (46.45)	0.791 (0.120)

Durations are in centiseconds.

Keystroke Duration

The statistical analysis of keystroke duration presents a special problem. Since all 21 subjects took part in all three window conditions, and each window condition contained three trials, the duration values might have been affected by practice effect. To test this possibility, a repeated measures multivariate analysis of variance (MANOVA) (SPSS-X, 1988) with windows and trials as repeated factors was performed. The results show that both windows ($F [40,2] = 4.88; p < .01$) and trials ($F = [40,2] = 8.43; p < .002$) differ significantly. There is, however, no interaction between windows and trials ($F [80, 4] = 1.20; p > .30$). Since there is no interaction, the data for windows and trials will be analyzed separately.

The significant difference among the trials suggests that the subjects displayed a practice effect. To determine the locus and the extent of the practice effect, the data for trials were collapsed across windows. Table 4 shows the group means and standard deviations for Trials 1 through 9. It is important to note that, because of the counterbalancing of experimental conditions among subjects, each trial contained an equal number of data points from each of the three window conditions. A repeated measures ANOVA shows that the nine trials differ significantly ($F [8, 180] = 6.88; p < .001$). A post-hoc Scheffe procedure shows that Trial 1 differs significantly from Trials 5, 6, 7, 8, and 9 ($p < .05$). Other pair-wise comparisons yield insignificant differences ($p < .05$).

TABLE 4: Group Mean Values for Keystroke Savings for Nine Trials Collapsed across Three Window Conditions

Trial	Mean	SD
1	271.04	71.29
2	240.93	47.22
3	226.42	35.95
4	225.26	36.67
5	209.29	48.13
6	200.08	50.06
7	204.22	39.17
8	196.61	51.67
9	199.52	50.63

It seems that the practice effect appears midway through the trials, regardless of the window sizes. In addition, the mean and standard deviation values shown in Table 4 suggest that most of the learning took place in the very first trial with small but incremental gains found in Trials 2 through 6. A lack of windows by trials interaction suggests that differences among windows cannot be attributed to the practice effect.

The mean duration per keystroke was 1.98, 2.18, and 2.44 seconds for window sizes 5, 10, and 15 words, respectively. A repeated measures ANOVA shows significant differences among means ($F [2,60] = 6.16; p < .003$). A post-hoc Scheffe multiple comparison shows that the means for 5- and 15-word and 10- and 15-word windows are significantly different ($p < .05$). The 5- and 10-word windows do not differ significantly for keystroke duration ($p > .50$).

DISCUSSION

Do different window sizes yield significantly different rates of communication? Mean letter and word durations, shown in Table 3, are not statistically significant across windows, indicating that the rate of communication is not affected by window size for the subjects employed in this study. The finding that the message preparation time in the 15-word window condition is equal to that in the 5-word window is somewhat surprising, considering that the rate of prediction is the highest and the number of keystrokes required is the lowest in the 15-word condition. Apparently, the increase in the visual search time for the target word in the larger window is responsible for the lack of difference in message preparation time between the two window conditions.

Standard deviation value (Table 3) is relatively large for Dur/L in the 15-word window, indicating that subjects took widely variable amounts of time to prepare messages of comparable length. Table 3 also shows that subjects took 46 centiseconds more per keystroke during the 15-word window condition as compared to the 5-word window. This means that, paradoxically, the keystroke duration *increased* significantly with higher predictions. These factors, taken together, suggest that searching the word selection menu in lexical prediction AAC programs places sig-

