Adaptive One-Switch Row-Column Scanning

Richard C. Simpson and Heidi Horstmann Koester

Abstract— **Row-column scanning is a very slow method of communication. Options for increasing text entry rate include 1) dynamically changing the configuration of the row-column matrix or 2) using rate-enhancement techniques like word prediction, but evidence suggests that increased cognitive load imposed by these methods on the user can result in little or no improvement in text generation rate. An alternative we are investigating is adapting a system's scan delay during run-time. Our goal is to allow a scanning system to adjust its parameters "on the fly" (as opposed to the current practice of setting parameters during clinical assessments). This paper describes the evolution of a one-switch row-column scanning system that adapts its scan rate based on measurements of user performance. Two experiments have been performed to explore the effects of automatically adapting scan delay on users' text entry rate. Our results indicate that automatic adaptation has the potential to enhance text-entry rate without increasing task complexity.**

Index Terms—**Augmentative communications, automatic adaptation, Bayesian networks, computer access, one-switch scanning.**

I. Introduction

ROW-COLUMN scanning is a technique used by individuals with severe disabilities for entering text and other data into computers and augmentative communication devices. It is an important method because it can be used with as little as one switch for input. A common implementation of row-column scanning with one switch requires three switch hits to make one selection from a two-dimensional (2-D) matrix of letters, numbers, symbols, words, or phrases, and is illustrated in Fig. 1. The first switch hit initiates a scan through the rows of the matrix. Each row of the matrix, beginning with the first, is highlighted in turn until the second switch hit is made to select a row. Each column of the row is then highlighted in turn until the target is highlighted, when the third switch hit is made to select the target. Variations on this theme are abundant and include column-row scanning and continuous row scanning which eliminates the first switch hit needed to initiate row scanning.

One-switch row-column scanning is a very slow method of communication. An able-bodied individual using an optimally-designed matrix of 26 letters and a space can produce between six and eight words/minute using this method [1], [2]. Despite its limitations, row-column scanning fills an important niche within technology access methods by providing an affordable alternative for individuals with limited movement and vocal abilities. Hence, despite increasing interest in speech recognition, eye-tracking, and direct-brain interfaces for accessing assistive technology, there remain valid reasons for seeking to enhance performance using row-column scanning.

A	B	C	D	E	F	G	H
I	J	K	L	M	N	O	P
Q	R	S	T	U	V	W	X
Y	Z						

(a)

A	B	C	D	E	F	G	H
I	J	K	L	M	N	O	P
Q	R	S	T	U	V	W	X
Y	Z						

(b)

A	B	C	D	E	F	G	H
I	J	K	L	M	N	O	P
Q	R	S	T	U	V	W	X
Y	Z						

(c)

Fig. 1. Three switch-hit row-column scanning. In (a), the system is row-scanning and the first row is highlighted. In (b), the target row has been reached and the switch has been pressed for the first time. In (c), the system is scanning through each column within the target row. The switch is pressed a second time when the target letter (U) is reached.

One method for increasing text entry rate with row-column scanning is to dynamically change the configuration of the matrix of items (letters, punctuation marks, words, etc.) to reduce the number of scan steps required to reach the most likely selections. Another alternative is the use of a rate-enhancement technique such as word prediction or abbreviation expansion. However, recent work indicates that the increased cognitive load these methods impose on the user may result in little to no improvement in text generation rate [3], [4].

Another alternative, which we are presently investigating, is the possibility of increasing a user's text entry rate by dynamically adapting the parameters controlling a system's scanning behavior during run-time. Table I lists the possible parameters that could be modified in an adaptive row-column scanning system. Our goal is to develop a method for matching these parameters to each individual user. If a given scan delay (row scan delay, column scan delay, initial row scan delay, initial column scan delay) is not long enough, the user's text entry rate is likely to decrease due to increased errors. A scan delay that is too long, on the other hand, will have few errors but will also fail to produce the maximum possible text entry rate.

Manuscript received April 17, 1998; revised August 23, 1999.

R. C. Simpson is with Texas Robotics and Automation Center (TRACLabs), Houston, TX 77058 USA.

H. H. Koester is with Koester Performance Research, Holland, MI 49424 USA.

Publisher Item Identifier S 1063-6528(99)09065-5.

TABLE I
ROW-COLUMN SCANNING PARAMETERS

Parameter	Explanation
Initial Row Scan Delay	Additional delay before the system begins scanning through rows
Initial Column Scan Delay	Additional delay before the system begins scanning through the columns within a row
Row Scan Delay	The amount of time a given row remains highlighted, the amount of time the user has to select the currently highlighted row
Column Scan Delay	The amount of time a given column remains highlighted, the amount of time the user has to select the currently highlighted column
Column Scans	The number of times the columns within a row are scanned if no selection is made once a row has been selected

The primary advantage of our approach is that it attempts to increase text entry rate without complicating the visual display. This implies that dynamically adapting the scan delay should impose less additional cognitive load upon the user than other rate enhancement techniques. Another advantage of this approach is that an adaptive system would be responsive to both permanent (e.g., due to learning) and transient (e.g., due to fatigue) changes in the user's ability.

Cronk and Schubert [5] represents an early attempt to automate the selection of row-column scanning parameters. An expert system was developed that, given a set of user characteristics, would usually make the same scan delay adjustments as a panel of three speech/language pathologists. The system was tested with six subjects, five of whom reported general agreement with the actions of the system. Unfortunately, no results were reported regarding the effect of the expert system's actions on user performance.

The approach taken in our research differs from this previous work in two respects. First, we are more interested in the effects that adaptation has on a user's performance than whether the adaptation decisions agree with a third-party expert. Second, the underlying reasoning method used to make decisions in our work is not a rule-based expert system but a probabilistic reasoning technique known as Bayesian networks [6]. Bayesian networks can be thought of as a means of organizing information to allow the convenient application of a form of Bayes' theorem

$$\Pr(H \mid e) = \frac{\Pr(e \mid H)\Pr(H)}{\Pr(e)}$$

where, in an adaptive row-column scanning system, H represents the choice of increasing, decreasing, or maintaining the current scan delay, e is a set of measurements of user performance, and the value of the probability $\Pr(H \mid e)$ represents the probability that performing a particular action (for example, increasing the scan delay) is the correct decision given the evidence presented to the Bayesian network. Because their reasoning is based on probabilistic information, Bayesian networks are well-suited for dealing with uncertain or conflicting information. An additional advantage is that a network's architecture provides insight into the nature and connections of the information sources being used to derive conclusions.

During the course of our research, several Bayesian networks have been designed and implemented within row-column scanning interfaces. This paper presents two experiments which evaluated the ability of individual networks to successfully adapt to the user characteristics of able-bodied subjects. Each experiment represents an incremental advance toward our ultimate goal of developing a robust system for adaptive one-switch row-column scanning that can be integrated with a variety of existing and future computer access and augmentative communication products.

II. EXPERIMENT 1

Experiment 1 compared the performance of two subject groups, one of which used an adaptive row-column scanning system, the other of which used a standared, nonadaptive row-column scanning system. The experiment focused on two key issues: 1) Would automatic adaptation make a difference in subjects' text entry rate and 2) would subjects agree with the decisions made by the system? Our belief was that automatic adaptation would result in greater text entry rate than manual adaptation. We also felt that subjects' ratings of the system's decisions would support the claim that the system was making reasonable decisions.

A. Testbed Row-Column Scanning Interface

An experimental one-switch row-column scanning system was developed specifically for this work, due to our need to integrate the Bayesian network directly into the row-column scanning interface and a desire for data collection capabilities not normally found in commercially available row-column scanning systems. The testbed row-column scanning system

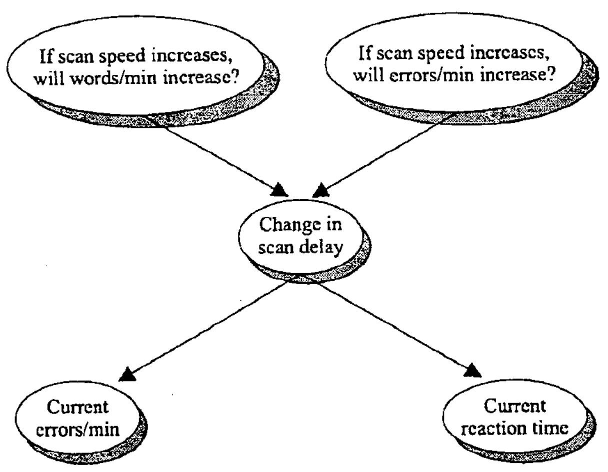

Fig. 2. Bayesian network used in Experiment 1.

presented subjects with a matrix of letters and punctuation marks, ordered by frequency of use. Three switch hits were required to make one selection from the matrix (see Fig. 1).

In the experimental system, there was one scan delay (defined as the amount of time a row or column remained highlighted before the next row or column became highlighted) for both rows and columns and no additional delays before the first row or column. The sentence the subject was expected to type was displayed beneath the selection matrix. As the subject entered text, the letters and punctuation that the subject had entered were shown beneath the target sentence.

B. Bayesian Network

Fig. 2 shows the Bayesian network used in this experiment. Each node in the network represents a probabilistic variable. The nodes at the bottom of the network (*Current errors/min* and *Current reaction time*) are directly observable variables whose values are determined by the user's performance over a given time interval (30 s for this experiment). The nodes at the top of the network represent prior probabilities that are determined by the system based on the user's performance during previous intervals. The arrows connecting nodes have both a causal and a probabilistic interpretation. The direction of the arrow between two nodes indicates that 1) the value of the node at the base of the arrow influences ("causes") the value of the node at the tip of the arrow and 2) the value of the node at the tip of the arrow is (probabilistically) dependent on the value of the node at the base of the arrow.

At the end of each time interval, the values of the upper and lower nodes are used to calculate the value of the middle node. The middle node corresponds to a probabilistic variable whose values range over the possible actions in response to the data observed during the current interval (lower nodes) and all previous intervals (upper nodes). The value of this variable determines 1) whether the system raises, lowers, or maintains the current scan delay and 2) how big a change (if any) is made. The process of recording data for an interval of time and then making an adaptation decision repeated continuously during use.

C. Method

Eight able-bodied subjects (two male, six female) ranging in age from 22 to 30 participated in an experimental evaluation of the performance of the adaptation expert. Subjects were randomly divided into two groups based on whether the adaptation expert did (condition AA) or did not (condition NA) automatically adapt the scan delay during trials. All subjects were given instructions as to the purpose of the experiment and the operation of the system and were then given the chance to practice entering text into the system before data was recorded.

Three measures of the subject's performance were recorded during the experiment: timing errors, missed errors, and adjusted text entry rate (ATER). *Timing errors* represent every occasion in which the subject pressed the switch at the wrong time (i.e., whenever they selected the wrong row or column). *Missed errors* represent every occasion in which the subject let the highlight scan past the target row or column. The *adjusted text entry rate* is calculated as the ratio of correct letter and punctuation selections (i.e., all correct column selections) to total time.

The structure of the experiment was the same for each subject regardless of experimental condition. Each subject had an initial scan delay of 750 ms. Each subject completed three one-sentence "warmups," with a rest period in between. Subjects then completed four four-sentence trials, with a warmup sentence in between each trial. After every trial and warmup the subject was given the opportunity to adjust the scan delay. Subjects were allowed to reduce the scan delay after a warmup or trial only if timing errors were less than 5% of total selections and the number of missed errors was less than 10% of total selections; otherwise they could only increase (or maintain) the current scan delay. The only difference between conditions was that, in the automatic adaptation (AA) condition, the Bayesian network also made adaptation decisions as the subject used the system. The Bayesian network was only active during trials (not during warmups) and decided after every 30 s in each trial whether or not to increment or decrement the scan delay. All adjustments (either up or down and either made by the subject or the Bayesian network) were made in 25, 50, or 75 ms increments.

After all trials were completed, subjects in the AA group were asked to fill out a questionnaire regarding their subjective impression of the Bayesian network's performance. All responses were given by placing marks on a line five inches long. At each end of the line for a question were vertical markers with phrases indicating extreme answers to the question being asked. Subjects' answers were converted to numerical scores between one and ten by measuring the distance of the subject's mark from the leftmost extreme of the scale. Table II lists each question and associated extreme values.

Two performance measures were compared between experimental conditions: total errors (the sum of timing errors and missed errors) and ATER. Two analyses were performed. First, data from all four trials was compared between conditions using an ANOVA with a between subjects factor of adaptation condition and a repeated measure factor of trial. Second, data from Trial 4 alone was compared between conditions using

TABLE II
QUESTIONS AND ASSOCIATED EXTREME ANSWERS FROM QUESTIONNAIRE GIVEN TO SUBJECTS IN GROUP AA

Question #	Question	Value of 1.0	Value of 10.0
1	How noticeable were the system's adaptations?	Not Noticeable At All	Very Noticeable
2	How often did you agree witht he adaptation decisions made by the system?	Never	Always

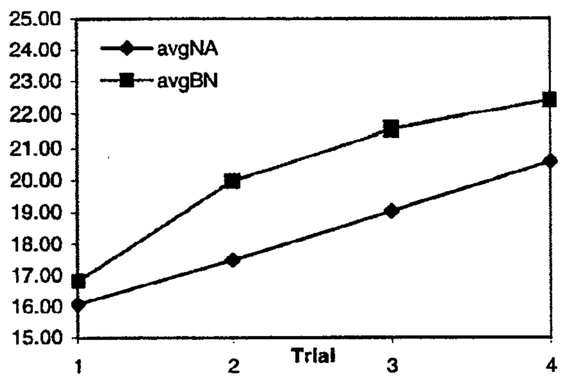

Fig. 3. Average ATER for Experiment 1.

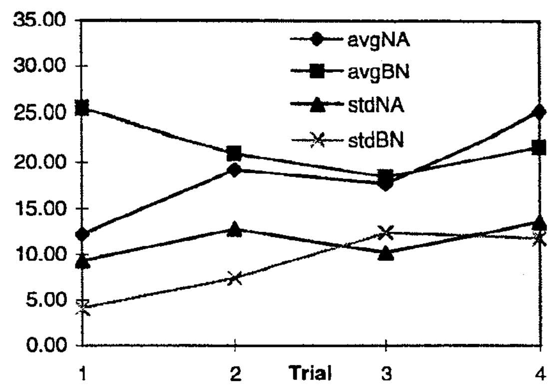

Fig. 4. Average total errors from Experiment 1.

a standard t-test. A significant difference was defined as a p-value less than 0.05.

D. Results

Fig. 3 shows the average ATER for both experimental conditions. As expected, ATER increased over trials as subjects gained proficiency with row-column scanning ($p < 0.0005$). In Trial 4, the average ATER for the AA condition was 22.41 characters per minute (cpm), with a standard deviation (SD) of 3.67, and the average for the NA condition was 20.59 cpm (SD = 4.88). The difference between conditions was not significant either over all four trials ($p = 0.44$) or for Trial 4 alone ($p = 0.57$).

Fig. 4 shows the average total errors for both experimental conditions. Interestingly, the average total errors for the AA condition dropped over the course of the experiment but rose for the NA condition. However, the effect of trials was not statistically significant ($p = 0.60$). In Trial 4, AA subjects made an average of 21.75 total errors (SD = 11.79), a number equal to 3.74% of total selections, and for NA it was 25.25 (SD = 13.60), equal to 4.28% of total selections. The difference in total errors between conditions, both over all four trials ($p = 0.60$) and for trial 4 alone ($p = 0.71$), was not statistically significant. Note that total errors include errors of ommission (failing to select the correct row or column when highlighted) so an error can occur without a corresponding selection.

Another intriguing result was the large split in average total errors between conditions for the first trial, which was statistically significant ($p < 0.05$). In Trial 1, the average total errors for the AA condition was 25.50 (SD = 4.12), equal to 4.72% of total selections, and for NA it was 12.25 (SD = 9.32), equal to 2.33% of total selections. One reason for this disparity was that subjects in group AA were scanning at much faster speeds than those in group NA by the end of Trial 1. Subjects in group NA had an average scan delay of 675 ms. Subjects in group AA, on the other hand, had an average starting scan delay of 644 ms and an average ending scan delay of 525 ms. As the scan delay for subjects in group NA caught up with the scan delay for subjects in group AA, their total errors converged.

In order to evaluate the quality of the decisions made by the Bayesian network, the actions of subjects and the Bayesian network must be separated. Table III shows the average decisions made by each. Subjects from both experimental groups changed the scan delay by approximately the same amount (−75 ms for group NA, −106 ms for group AA) during the warmup period before Trial 1. The Bayesian network reduced the scan delay by an average of 25 ms per trial for subjects in the AA group. This made up slightly less than half of the total average adjustment made per trial by group AA. The total adjustment made per trial (the sum of average change during a trial and average change following that trial) for the two experimental conditions was almost equal (−52 ms for group NA, −60 ms for group AA). This similarity in total timing adjustments per trial is a main reason why ATER did not differ significantly for the two groups.

An illustration of how the Bayesian network adjusted the scan delay during a single trial is given in Fig. 5, which shows

TABLE III
SCAN DELAY ADJUSTMENTS MADE BY THE HUMAN SUBJECT AND ADAPTATION EXPERTS

Group	Average Change in Scan Delay Before Trial 1	Average Change During Trials by Adaptation Experts	Average Change Between Trials by Subjects	Average Total Change Between and During Trials
NA	-75.00 ms	0.00 ms	-52.08 ms	-52.08 ms
AA	-106.25 ms	-25.00 ms	-35.42 ms	-60.42 ms

data from one trial for one subject in group AA. As can be seen from the graph, the Bayesian network often increased the scan delay at the end of 30-second intervals in which the number of errors rose and decreased the scan delay after intervals in which no errors occurred. However, because the subject's reaction time and total errors, as well as past performance measures, were all taken into account to decide what action should be taken after that interval (see Fig. 2), neither response was guaranteed.

As described above, each subject in group AA was asked to rate how often they agreed with the adaptation decisions made by the system from one (never) to ten (always). The average score was 7.47 (95% confidence interval of [6.06, 8.88]). Subjects were also asked to rate how noticeable the system's actions were from one (not at all) to ten (very). The average score was 5.64 (95% confidence interval of [3.02, 8.26]).

E. Discussion

Two measures of the Bayesian network's performance were of interest in this experiment, the first being the network's impact on subjects' text entry rate. There was not a significant difference in text entry rate between the two experimental conditions, which indicates that the subjects within the NA group were quite adept at accurately determining their own best scan delay. This implies that the Bayesian network is unlikely to improve upon the performance of an able-bodied user in a system with a single scan delay, but is capable of matching the performance of an able-bodied user. It is also reasonable to expect that the Bayesian network would improve performance in a system with multiple parameters that can all be set independently.

The decisions made by subjects indicated that the adaptation expert was not aggressive enough in terms of adjustments to the scan delay. The Bayesian network shared the parameter adjustment task with test subjects, which reduced, but did not eliminate, their need to make manual adjustments on their own. The adaptation expert often made the correct decision in terms of which direction to adapt the scan delay, as indicated by subjective rankings and lack of difference in text entry rate or errors, but did not make large enough adjustments, as indicated by the large average between trial adjustment by members of the AA group.

The second measure of interest was the extent to which the Bayesian network's actions met users' expectations. Space does not permit a detailed presentation of the individual decisions made by the Bayesian network during each trial, but subjects' ratings of the Bayesian network's decision-making supported the conclusion that the Bayesian network generally made reasonable decisions given the data it was confronted with in each trial.

The results of Experiment 1 left several questions unanswered. In particular, it was unclear whether performance between groups would have been as similar if trials had been longer or subjects had not been given so many opportunities to manually adjust the scan rate. In order to address some of the perceived shortcomings in the first experiment, and to determine whether the performance of the Bayesian network could be improved, the Bayesian network was redesigned and evaluated in a second experiment.

III. EXPERIMENT 2

Our goals for this experiment were 1) to demonstrate that the Bayesian network could entirely assume the task of adjusting the scan delay (i.e., without any user intervention at all) and 2) to determine subjects' perceptions about the performance of the system. In pursuit of these goals, several changes were made to the protocol used for Experiment 1, and the adaptation expert was modified to improve its performance. The revised experimental protocol allowed subjects to directly compare both manual and automatic adaptation. Subjects completed fewer trials than the previous experiment (two), but each trial was longer to allow for the accumulation of many adaptation decisions during a single trial. All subjects in this experiment completed a trial with automatic adaptation active and a trial without automatic adaptation. The changes to the adaptation expert made it more aggressive and expanded the information it used to make decisions.

A. Testbed Row-Column Scanning Interface

A new row-column scanning testbed was implemented for Experiment 2. A primary motivation for this was to shift the software from the MS-DOS operating system to the Windows95 operating system, to support further studies which will take place in the Windows environment. The scanning interface was similar to the interface used in Experiment 1,

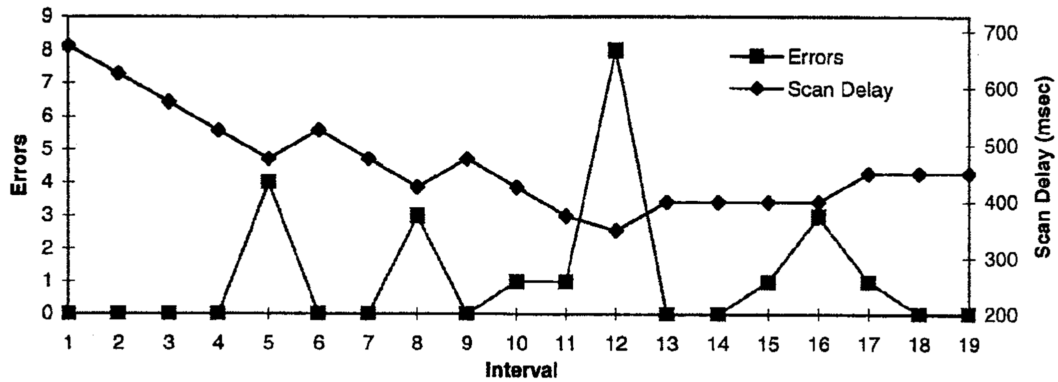

Fig. 5. Errors and scan delay from one trial in which the Bayesian network was active.

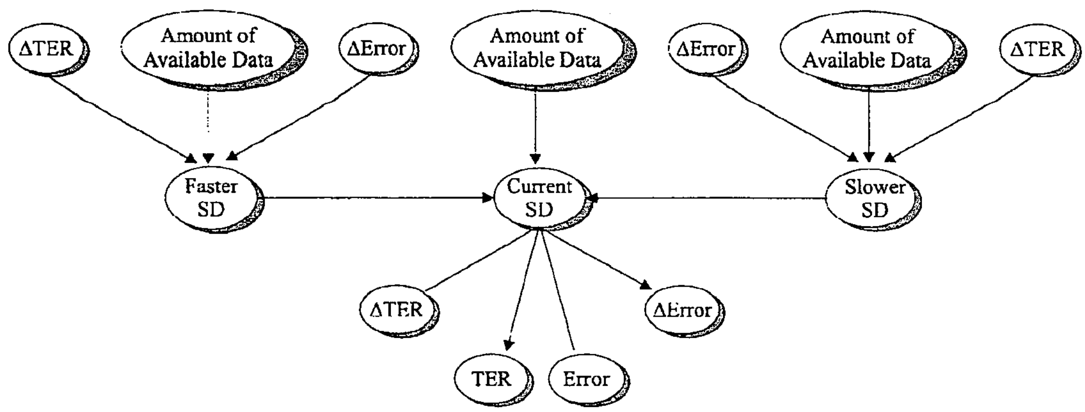

Fig. 6. Bayesian network used in experiment 2. TER = Text entry rate, ΔTER = Change in text entry rate, Error = Number of errors, ΔError = Change in number of errors, and SD = Scan delay.

except that subjects did not have to press the switch to initiate row-column scanning, which meant that only two switch hits (one for the correct row, one for the correct column) were required to select a target from the matrix.

Like the first experimental system, there was one scan delay for row and column scanning, with no additional delay before the first row or column. The target sentence was displayed beneath the selection matrix, and the user's input was shown beneath the target sentence. Unlike the previous experimental system, the new interface provided both auditory and visual cues when an adaptation decision was made. The system beeped whenever the user chose to change the scan delay or whenever the adaptation expert made a decision (either to change the scan delay or maintain the current scan delay). The visual cues consisted of a text box, displayed on the screen, which reported the last adaptation decision made by the adaptation expert. These changes were made to increase subjects' awareness of the adaptation decisions made by the Bayesian network.

B. Bayesian Network

The Bayesian network used in Experiment 2 is shown in Fig. 6. The decision-making algorithm represented by the network is a comparison of the user's performance at the current scan rate to the user's performance at the next faster and next slower scan rate. Likely performance at faster and slower scan rates is estimated based on 1) the average change in text entry rate and errors observed at the faster (or slower) scan rate versus the current scan rate and 2) the amount of data these averages are based on (in terms of number of decision intervals at that scan rate). Future performance at the current scan rate is projected based on the current text entry and error rate and trend of these averages. The decision made by the network is based on the resulting probability that performance will be greatest using the current, next-faster, or next-slower scan delay.

The Bayesian networks shown in Figs. 2 and 6 have several key differences. While both networks consider the same basic information, the observed and projected errors and text entry rate at different scan rates, they integrate this information in different ways. In particular, the network used in the second experiment uses the available information to a greater extent by calculating trends in total errors (the ΔError node) and text entry rate (the ΔTER node), keeping track of how much information has been accumulated at different scan rates (the Amount of Available Data nodes), and using estimates of performance results from both faster and slower scan rates (the Faster SD and slower SD nodes) in its decision making process.

When the adaptation expert was active and the subject was entering text, the system recorded performance data for a predetermined time interval (10 s in this experiment). At the

TABLE IV
QUESTIONS AND ASSOCIATED EXTREME ANSWERS FROM QUESTIONNAIRE GIVEN TO EACH SUBJECT

Question #	Question	Leftmost Extreme	Rightmost Extreme
1	How difficult was the task under manual control?	Very Difficult	Very Easy
2	How difficult was the task under automatic control?	Very Difficult	Very Easy
3	How often did you agree with the system's decisions under automatic control?	Never	Every Time
4	How would you rate your decisions making during manual control?	Very Bad	Excellent
5	How would you rate the computer's decision making during automatic control?	Very Bad	Excellent
6	How noticeable were changes in scan delay under automatic control?	Very Noticeable	Not Noticeable At All
7	What effect did manual control have on your performance?	Negative	Positive
8	What effect did automatic control have on your performance?	Negative	Positive
9	Which condition did you prefer?	Automatic Control	Manual Control

end of each interval, the adaptation expert used the recorded information to decide whether to raise, lower or maintain the current scan delay. However, in this experiment the adaptation expert was limited to changes of 25 ms in the scan delay.

C. Method

Eight able-bodied subjects (six male, two female) ranging in age from 23 to 56 participated in an experimental evaluation of the performance of the adaptation expert. No subjects from Experiment 1 participated in Experiment 2. Subjects were randomly divided into two groups based on whether the adaptation expert (Group 1) or the subject (Group 2) was in charge of adaptation decisions in the first trial. All subjects were oriented as to the purpose of the trial and the operation of the system and practiced entering text into the system before data was recorded.

The structure of the experiment was the same for each subject. In every training session and trial, the initial scan delay was set to 750 ms. Each subject completed a two-sentence training session with the system configured for manual or automatic adaptation depending on the subject's group membership. Subjects then entered text for a ten sentence trial in which every keystroke, error, and adaptation decision was recorded and time-stamped. Subjects then completed a second training session and ten-sentence trial in which the system was configured for the opposite experimental condition. All the sentences used for training sessions and trials were divided into two-sentence blocks. The number of scan steps (movements between rows or columns by the highlight bar) were matched across all blocks to allow results to be validly compared between blocks. The blocks served as the unit of statistical analysis, rather than data being averaged over an entire trial. This allowed data to be analyzed over time and increased the statistical power of the analysis.

When automatic adaptation was active only the adaptation expert could make adaptation decisions. Similarly, only the subject could make adaptation decisions under the manual adaptation condition. Both the adaptation expert and the subject were restricted to changing the scan delay in 25 ms increments. As described above, the adaptation expert made an

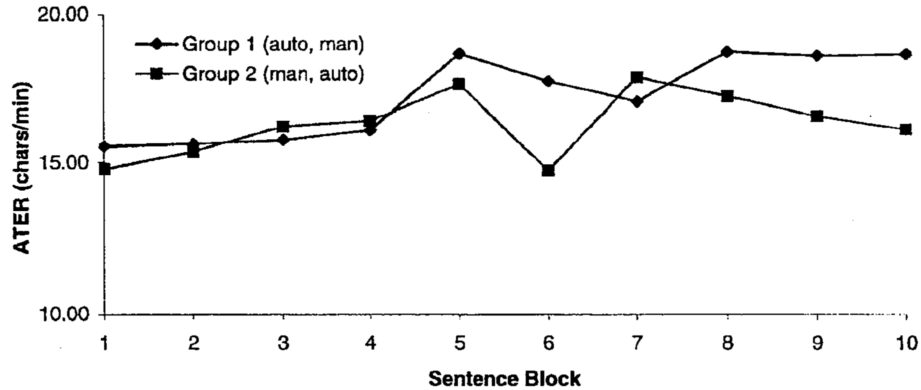

Fig. 7. Average ATER for Experiment 2.

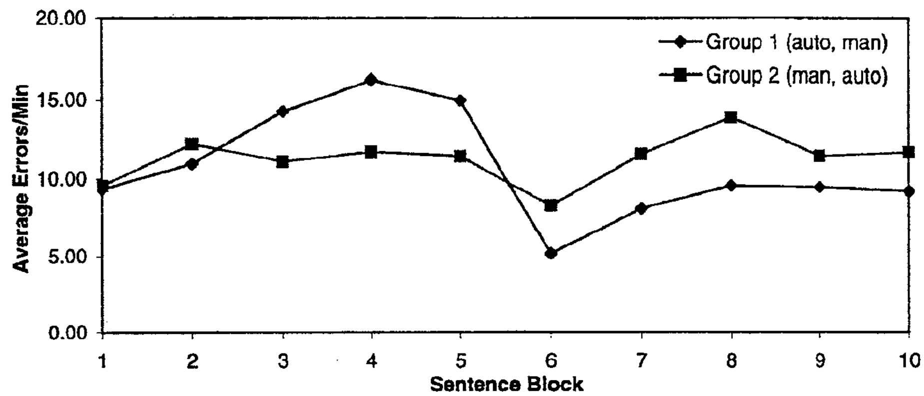

Fig. 8. Average errors per minute for Experiment 2.

adaptation decision every 10 s. Under the manual condition, however, the subject had no limits on the number or frequency of adaptation decisions. Subjects changed the scan delay by pressing the up arrow key to scan faster (decrease the scan delay) and pressing the down arrow key to scan slower (increase the scan delay). Unlike Experiment 1, subjects were able to manually adjust the scan delay during both warmups and the trial under the manual adaptation condition.

After both trials were completed, subjects were asked to fill out a questionnaire on their subjective impression of each condition. All responses were given by placing marks on a line four inches long. At each end of the line for a question were vertical markers with phrases indicating extreme answers to the question being asked. Subjects' answers were converted to numerical scores between one and five by measuring the distance of the subject's mark from the leftmost extreme of the scale (which corresponded to an answer of 1.0). A mark on the rightmost extreme corresponded to an answer of 5.0 and a neutral answer (placed exactly between the two extremes) corresponded to an answer of 3.0. Table IV lists each question and associated extreme.

Adjusted text entry rate (ATER) and total errors were compared between conditions based on averages over each two-sentence block. Responses to the questionnaire were compared between groups and to the neutral answer (3.0). Several statistical analyzes were performed, as described below, with a significant difference defined as a p-value less than 0.05.

D. Results

Fig. 7 shows the average ATER for both subject groups over both trials. Sentence Blocks 1 through 5 correspond to the ten sentences of the first trial while sentence Blocks 6–10 correspond to the ten sentences of the second trial. Text entry rates for each subject group were compared over Blocks 1–5, and then again across Blocks 6–10, using a repeated measures ANOVA with a between-subjects factor of adaptation condition and a within-subjects factor of sentence block. For Blocks 1–5, adaptation condition (automatic or manual) was not statistically significant ($p = 0.912$) while block was ($p = 0.018$). For Blocks 6–10, adaptation condition was not significant ($p = 0.508$) nor was block ($p = 0.312$). This implies that adaptation did not have an effect on ATER in either trial and that practice did not have an effect on performance in Trial 2. However, performance does seem to have improved as subjects gained experience during Trial 1.

Fig. 8 shows the average errors per minute for both subject groups over both trials. As above, errors for each group were compared using repeated-measures ANOVA's over Blocks 1–5 and 6–10. For Blocks 1–5, adaptation condition was not significant ($p = 0.553$) and block was not significant ($p = 0.082$). For Blocks 6–10, adaptation condition was not significant ($p = 0.260$) but block was significant ($p = 0.049$). This means that adaptation condition did not have an effect on total errors in either trial and that experience did not have

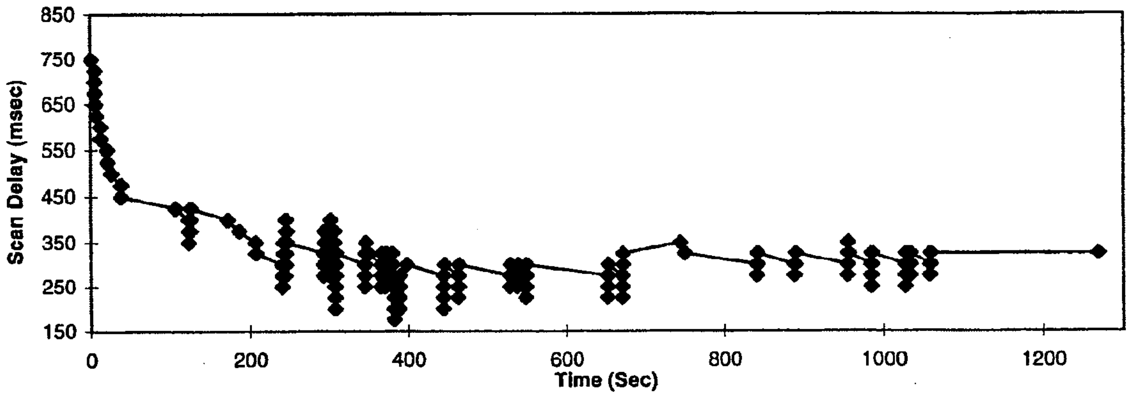

Fig. 9. Subject MM manual adaptation.

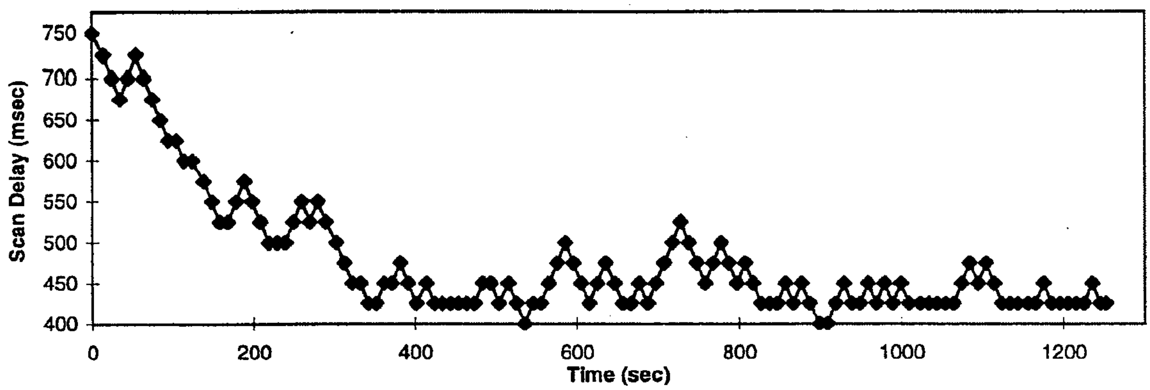

Fig. 10. Subject MM automatic adaptation.

an effect on error in Trial 1. Interestingly, practice did have an effect on errors in Trial 2, but (paradoxically) fewer errors were made in early blocks than later blocks.

Table V shows the results from subjects' responses to the questionnaire. There was not a significant difference between the total average response for any question and the neutral answer (3.0). The subjective data also did not demonstrate a significant difference for any question based on adaptation condition.

Fig. 9 shows the scan delay over time from a single subject from group 2 (the group that used manual adaptation for trial 1) during a trial in which manual adaptation was active. This can be compared to Fig. 10, which shows the scan dèlay for the same subject during the trial in which automatic adaptation was active. As can be seen from the figures, when subjects were in charge of changing the scan delay, they tended to perform changes in bursts with distinct pauses in between. When automatic adaptation was active, on the other hand, there was at least ten seconds between each change in scan delay.

TABLE V
AVERAGES OF RESPONSES TO THE QUESTIONNAIRE. 1.0 WAS THE LEFTMOST EXTREME, 3.0 WAS THE NEUTRAL ANSWER AND, 5.0 WAS THE RIGHTMOST EXTREME

Question #	a/m Grp 1 Avg	m/a Grp 2 Avg	Total Avg
1	3.67	3.70	3.69
2	2.67	3.56	3.12
3	2.69	3.41	3.05
4	3.83	3.31	3.57
5	2.52	3.23	2.88
6	2.45	3.17	2.81
7	4.39	3.22	3.80
8	3.00	3.70	3.35
9	4.34	2.09	3.22

E. Discussion

As stated above, one of our goals was to develop an adaptation expert which would be able to completely take over the task of setting the scan delay. The results indicate that using the adaptation expert to control the scan delay did not cause a significant difference in text entry rate from an experimental condition in which subjects had complete control over scan delay. In addition, there was not a statistically significant difference between conditions for total errors. However, the data suggests that some subjects made slightly more errors when automatic adaptation was active. This difference in errors will need to be examined further in future studies.

One possible explanation for the increased errors is people's preference for developing a scanning "rhythm," which automatic adaptation interfered with. Several subjects reported that their best performance came when the scan delay remained fixed at a given speed long enough for them to synchronize their actions to the scan delay. When automatic adaptation was active, the scan delay could change every ten seconds, which made synchronization more difficult. This does not mean that automatic adaptation is inherently flawed. However, it does indicate that changes should probably happen less frequently than they occurred during the experiment and that the notion of preserving the subject's rhythm should be incorporated into future enhancements of the adaptation expert. One possible mechanism for accomplishing this is to only make changes to the scan delay between sentences.

Our other stated goal was to determine subjects' feelings about both automatic and manual adaptation. There was very little difference between the neutral answer and the average response for most questions, and no question showed a statistically significant difference based on group. This implies that subjects did not markedly prefer manual or automatic adaptation. However, there was a noticeable and consistent difference between groups for question 9, which asked which condition subjects preferred. As a group, subjects in Group 1 (the group that used automatic adaptation for Trial 1) preferred automatic adaptation while subjects in Group 2 preferred manual adaptation. One reasonable explanation for this result is that the added cost of relearning the task in Trial 2 predisposed subjects to prefer the first experimental condition they experienced.

It is important to realize that the manual adaptation condition, as implemented within this experiment, represents an idealized situation that does not correspond to the typical row-column scanning scenario. First, we are unaware of any commercially-available row-column scanning system that makes changes in the scan delay as directly accessible as a single keystroke. In addition, the user of our manually adaptable row-column scanning interface needs to be able to consistently use three switches: one to make selections and two to change the scan delay. In reality, the average one-switch row-column scanning user is using a one-switch system because they find it difficult to operate multiswitch input methods. Hence, actual users of row-column scanning are unlikely to find the option of direct manipulation of the scan delay as useful as the subjects that participated in the experiment.

IV. Conclusion

The results achieved in our experiments are promising, but the protocols used in each experiment had limitations. A primary limitation was the lack of disabled subjects. In these preliminary experiments it was simply more practical to use able-bodied subjects. Another limitation of our experimentations was the short amount of time that each subject used the row-column scanning system. Once again, expediency dictated the format of these preliminary experiments. Future work is planned to involve subjects with disabilities and to expand the amount of time each subject interacts with the system.

Our long-term goal is to develop a row-column scanning system that can adapt several interface settings (row scan delay, column scan delay, initial row scan delay, initial column scan delay, number of times the columns within a row are scanned if no selection is made) without requiring any assistance from the user. The results of these preliminary experiments indicate that we have achieved an important preliminary milestone by demonstrating that an appropriately designed Bayesian network can make reasonable adaptation decisions (with no human intervention) for a system with a single scan delay.

Of equal importance, the presence of automatic adaptation neither hindered nor enhanced subject performance. Our belief is that the user is the best judge of what the most appropriate scan delay should be. Hence, we have used the performance of the user under the manual adaptation condition as the "gold" standard to which we compare the performance of the system under the automatic adaptation condition. However, we also expect performance of a user to be adversely affected as the number of interface parameters that they are asked to monitor and manually adapt increases. Future work is planned to determine whether performance is, in fact, improved when the system can change multiple system parameters during run-time.

References

[1] R. Damper, "Text composition by the physically disabled: A rate prediction model for scanning input," *Appl. Ergonom.*, vol. 12, pp. 289–296, 1984.

[2] H. Koester and S. Levine, "Learning and performance of able-bodied individuals using scanning systems with and without word prediction," *Assist. Technol.*, vol. 6, no. 1, pp. 42–53, 1994.

[3] P. Demasco, "Human factors considerations in the design of language interfaces in AAC," *Assist. Technol.*, vol. 6, no. 1, pp. 10–25, 1994.

[4] H. Koester, "User performance with augmentative communication systems: Measurements and models," Ph.D. dissertation, Univ. of Michigan, Ann Arbor, 1994.

[5] S. Cronk and R. Schubert, "Development of a real time expert system for automatic adaptation of scanning rates," in *Proc. 10th Annu. Conf. Rehab. Technol.*, Washington, DC, RESNA, 1987, pp. 109–111.

[6] E. Charniak, "Bayesian networks without tears," *AI Mag.*, vol. 12, no. 4, pp. 50–63, 1991.

Richard C. Simpson, for a photograph and biography, see this issue, p. 451.

Heidi Horstmann Koester received the Sc.B. degree in biomedical engineering from Brown University, Providence, RI, in 1986, and the M.S. degree in bioengineering, the M.A. degree in psychology, and the Ph.D. degree in bioengineering all from the University of Michigan, Ann Arbor, in 1988, 1993, and 1994, respectively.

She is currently the President of Koester Performance Research, an assistive technology research and consulting firm. Her educational training includes work in human performance, human–computer interaction, software engineering, human disability, experimental research methods, and statistical analysis. She has years of clinical experience providing computer access and augmentative communication services to children and adults with disabilities. Her research focuses on user performance with assistive technology systems, from both theoretical and clinical perspectives. A theme running throughout the work is the application of human–computer interaction theory and quantitative methods to practical clinical issues. Publications resulting from this work have won several international awards. The research employs a variety of methods including analytical methods of user interface analysis, user surveys, traditional group designs, and single-case designs.