Multimodal Access to Written Communication

by

Jutta Treviranus

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Department of Human Development and Applied Psychology
Ontario Institute for Studies in Education of the University of Toronto

© Copyright by Jutta Treviranus, 1996
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non-exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/filmm, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-28724-6
Multimodal Access to Written Communication
Master of Arts, 1996
Jutta Treviranus
Department of Human Development and Applied Psychology
Ontario Institute of Studies in Education of the University of Toronto

Abstract

Many children and young adults who are severely physically disabled and functionally non-speaking use scanning techniques to control computers. Access to written and face-to-face communication using scanning is frequently far too slow to meet functional requirements. This study investigated whether combining speech recognition with scanning can increase the rate of computer input for these individuals. The study compared two access techniques: traditional scanning and scanning combined with speech recognition. Analysis of the performance of eight children and young adults, who employed scanning as a computer access means, was used to determine the relative input rate and accuracy of the two techniques. A significant increase in items entered per minute was found when comparing access through scanning combined with voice recognition versus access through scanning alone. No significant differences in accuracy were evident.
Acknowledgements

This project was partially funded by The Easter Seal Research Foundation, The Hospital for Sick Children Foundation, and Apple Canada Ltd. I would like to acknowledge the staff of the Microcomputer Applications Programme for their technical assistance. I would also like to thank the individuals who participated in the study for their patience and willingness to try new things.
# Table of Contents

Abstract.................................................................................................................. ii
Acknowledgments....................................................................................................... iii
Table of Contents........................................................................................................ iv
List of Tables................................................................................................................ vi
List of Figures............................................................................................................... vii

Chapter 1: Introduction................................................................................................. 1

Chapter 2: Background................................................................................................. 3
  Computer Input Systems......................................................................................... 3
    Direct Selection................................................................................................. 3
    Scanning ........................................................................................................... 4
  Acceleration Techniques .................................................................................... 4
  Expanding the Communication Channel ......................................................... 5
  Bridging the Gap between Direct Selection and Scanning ................................. 5
  Human Factors Considerations in Designing the Interface ............................... 6
  Speech Recognition Systems............................................................................... 6
  Combining Speech Recognition with Scanning................................................... 7

Chapter 3: Method ...................................................................................................... 8
  Experimental Condition ...................................................................................... 8
    Participants ...................................................................................................... 10
    Speech Screening ............................................................................................. 11
    Speech Recognition Screening ........................................................................ 11
    Speech Recognition System ........................................................................... 12
    Testing Software ............................................................................................. 12
  Procedure ............................................................................................................ 13
    Introductory Sessions ...................................................................................... 13
    Learning Sessions ........................................................................................... 14
    Decision Criteria ............................................................................................. 15
    Control for Order Effect .................................................................................. 15
  Measures ............................................................................................................. 16
    Copy Type Tasks ............................................................................................ 16
    Error Classification ......................................................................................... 16
    Subjective Ease of Use and Preference Ranking .............................................. 16

Chapter 4: Results ..................................................................................................... 18
  Order Effect ........................................................................................................ 18
List of Tables

Table 1: Participant profiles ................................................................. 11
Table 2: Number of items entered per minute and errors .......................... 18
Table 3: Subjective ease of use and preference ranking ............................ 22
List of Figures

Figure 1: Sample personalized visual keyboard .............................................. 8
Figure 2: Participant 1, items per minute, scan and voice and scanning alone .......... 19
Figure 3: Participant 2, items per minute, scan and voice and scanning alone .......... 19
Figure 4: Participant 3, items per minute, scan and voice and scanning alone .......... 19
Figure 5: Participant 4, items per minute, scan and voice and scanning alone .......... 20
Figure 6: Participant 5, items per minute, scan and voice and scanning alone .......... 20
Figure 7: Participant 6, items per minute, scan and voice and scanning alone .......... 21
Figure 8: Participant 7, items per minute, scan and voice and scanning alone .......... 21
Figure 9: Participant 8, items per minute, scan and voice and scanning alone .......... 21
Chapter 1: Introduction

Although children with severe physical disabilities have access to alternative writing systems (e.g., scanning, Morse code) these are often slow and restrictive (Vanderheiden and Kelso, 1987). These children are therefore greatly handicapped in their attempts to keep up with peers in integrated educational settings. Strategies to accelerate access methods (e.g., abbreviation-expansion, prediction) have made relatively minor gains at the expense of additional cognitive processing demands (Light, 1989). Tapping a larger proportion of the child’s abilities through a "hybrid" access method, which combines two or more methods of controlling the computer, has not yet been investigated.

Instead of trying to accelerate input using a single access technique, it may be more effective to combine different methods into hybrid techniques. Many children with disabilities who converse by selecting items from communication displays which are not computer based, use a combination of scanning and direct pointing. Although their pointing skills are insufficient to pick out individual items, these children can use eyegaze or gross pointing to indicate groups of items. Listeners can then call out or point to items in a group until the child signals that the desired item has been reached. Since the child directly points for speedy access to groups of items, and need only scan within a group, less scanning is required to select an item.

Although they may not have sufficient physical control even for gross pointing, many children who are nonspeaking are able to vocalize. The average listener may not identify these vocalizations as words, but the utterances may be distinguishable from one another. It may be possible, therefore, to make use of these utterances, supplementing the scanning systems which these children generally use with a speech recognition system which maps words or word approximations into computer input. Distinguishable utterances could serve as direct selectors, performing the pointing function which these children cannot physically accomplish. In a hybrid system, children could either scan through or directly select groups of items, followed by items within the groups.

Some commercially-available speech recognition systems do not require intelligible words; they simply match a voice pattern with one previously recorded. Fried-Oken (1985), Schmitt and Tobias (1986), and Rodman et al. (1984) report the successful use of
such speech recognition systems with dysarthric speakers to write, to control their environment, and to clarify speech. At the Hugh MacMillan Rehabilitation Centre in Toronto, clinical observations of the use of speech recognition with several children and adults who are functionally nonspeaking but who are able to make a number of distinct utterances have noted a recognition accuracy of 70 to 88 percent with a vocabulary of 12 to 20 utterances. Although speech recognition would not be feasible as the sole input method for these individuals, it could accompany other input methods as part of a hybrid selection technique.

An important consideration in employing multiple input modes, especially with children, is cognitive processing. Forren and Mitchell (1986) found that combining speech recognition with keyboard entry slowed down able-bodied typists. They postulated that this was due to increased demands on the user's attention and processing resources. The user was required to devote cognitive resources to speaking each word in isolation, which is an unnatural method of speech, as well as attending to visual feedback in order to confirm entries.

For users of scanning systems, directly selecting by vocalization rather than passively waiting for the computer to present choices may in fact reduce the perceived cognitive load. Forren and Mitchell's conclusions may not apply to these users, since they are accustomed to dividing their visual attention between the text they are editing and the feedback from their scanning systems. In a hybrid access system, voice recognition feedback could be incorporated into scanning feedback in order to minimize additional cognitive processing demands.

Theoretically, combining speech recognition with scanning could provide a faster, more direct means of computer access for people who use scanning. It is important to investigate the effect on rate, accuracy, and cognitive load for nonspeaking individuals.

This report describes the results and conclusions of a two year study investigating the effects of combining speech recognition with scanning to access computers, for individuals with very limited speech. Specifications for more appropriate voice recognition technology are discussed. Strategies for implementing voice recognition with users who have limited speech are recommended.
Chapter 2: Background

Successful participation in conversations, and educational and vocational activities usually requires communication rates, timing and accuracy within certain boundaries of tolerance (Kraat, 1987). Unfortunately individuals who use alternative computer input systems speak at 1% to 5% and write at 10% of the rate of individuals using traditional means (Foulds, 1980; Vanderheiden, 1984). Frequently the operational demands of these input systems detract attention from the task itself (Light, 1988). The major problems with access systems available to individuals with severe physical impairments can be summarized as follows:

- they are too slow,
- operation of the access techniques is too demanding cognitively, and
- the access technique is required before the child has developed the requisite skills to operate the technique (Kraat, 1987; Light, 1990).

Computer Input Systems

Computer input systems can be roughly divided into two categories: direct selection where the user points to the desired item (or picks it out by naming it), or scanning where the user chooses from items sequentially presented by the device or partner through some predetermined signal (Vanderheiden, 1984).

Direct Selection

The simplest and fastest form of selection is pointing or direct selection. Pointing ability is defined by the area that can be covered with a pointing limb (range) and by the number of discrete targets that can be selected within that range (resolution). If the number of items the user requires to communicate or control an application closely matches the number of targets they can directly point to they can use a direct selection technique. The user need not point to a standard computer keyboard or use their fingers to point. The standard keyboard can be adapted (e.g., key delay to reduce unintentional key repeats, keyguards to reduce unintentional key hits), alternative keyboards can be used (e.g., keyboards with enlarged or miniaturized keys), and the user could point with other body parts (e.g., eyegaze, feet) (Lee, 1990).
Scanning

Many AAC users can only point to or make contact with one or two discrete targets. Such users must use scanning to select the elements of their intended message. Scanning is the sequential presentation of choices (letters, words, or pictures). A person or a computer presents the options, and the user signals when the desired item has been reached. In computer-based scanning, the options are usually highlighted by a cursor moving in an established pattern over an array of items on the computer screen. The user signals by activating or releasing one or more switches. Switches are carefully selected and positioned for operation with physical movements which the user can reliably control (Lee, 1990; Treviranus, 1990). Large sets of items are often arranged into subgroups; the user first chooses the group the desired item belongs to and then the item itself. Scanning is much slower and more demanding cognitively than direct selection (Blackstone, 1989; Light, 1988; Vanderheiden, 1984).

Users who have three to five voluntary movements which they can sustain and accurately release may be able to use directed scanning, where a cursor is driven across the matrix of selectable items to the intended item and a selection is indicated. This method may provide greater speed and directness than traditional scanning (Treviranus, 1990; Vanderheiden, 1984).

Acceleration Techniques

Several acceleration techniques have been developed which allow item selections to convey larger quantities of information, thereby increasing the rate of input of slower input techniques such as scanning. Abbreviation expansion is an acceleration technique that translates one or a few selections into a longer message which would otherwise require many selections (Light, 1989; Vanderheiden, 1987). A second technique known as linguistic prediction can complete the spelling of a word once a few letters have been entered (Swiffen et al., 1988). Although these techniques can significantly decrease the number of keystrokes or selections required to enter a given amount of text, researchers have noted that the cognitive and perceptual processes involved in using these techniques may counteract the gains in selection efficiency (Horstman and Levine, 1991).
Expanding the Communication Channel

Traditional scanning is slow and complex primarily due to the fact that all information is funneled, as it were, through one or two signals or communicative acts. The communication channel is severely constrained (Shein et. al, 1990). The one or two communicative signals can select the hundred or potentially thousands of possible messages using timing. By employing timing as a method of differentiating selections the process of indicating a choice becomes a lengthy process: many unwanted choices must be passed prior to choosing the desired item. Abbreviation-expansion or prediction can be used to reduce the number of scanning steps or the number of selections to be eliminated, but at the expense of additional cognitive load. Any addition of communicative signals, thereby widening the input channel, would proportionately reduce the number of scanning steps required to make selections. Additional communicative signals could be used in a number of ways:

• to modify the scanning strategy e.g., from automatic scanning to directed scanning
• to directly select a group of items which could then be scanned
• to directly select a high frequency item or function

Any of these modifications to the access system potentially improves speed and efficiency.

Bridging the Gap between Direct Selection and Scanning

The majority of commercially available alternative access systems employ either direct selection or scanning. As a result additional voluntary actions or communicative acts a user can produce are not exploited to control the access system. Three approaches can be used to bridge the gap between direct selection and scanning and exploit these additional communicative acts. One approach is to combine direct selection and scanning in one device. Thus, users would directly select to the extent of their abilities, and then scan to choose items which they are unable to select, or select from a larger group and then scan items within that group (hybrid selection). Hybrid selection techniques are most frequently used by individuals who communicate using graphic displays (Trefler & Crislip, 1985). Communication displays are divided into blocks which the user is able to select directly. Once the block is indicated the listener scans through the block waiting for the appropriate signal from the AAC user. Another approach is to use a form of coding where two or more direct selections are used to choose each item. The sequence of actions required to choose an item could be either cued or memorized. In the third
approach the user chooses a group of items and relies upon the computer to guess (or disambiguate) which item within the group is intended. Thus the user makes only one direct selection to choose each item, unless the computer has “guessed” wrong. Hybrid selection (combined with a simple form of visually cued coded access) were explored in this research project.

**Human Factors Considerations in Designing the Interface**

The design, method of introduction and training strategy used are critical to the successful application of an alternative access system. In designing a hybrid alternative computer access system the following considerations are relevant. Auditory or visual vigilance to external prompts or cues should be kept to a minimum. Even though two modes of input are used, voice and switch activation, feedback acknowledging the user’s input should be unified. The number and variety of steps required to complete the task should be kept to a minimum. Decisions to be made should be minimized and wherever possible the decisions should be routinized, or unambiguous criteria for making decisions should be provided. The user should be assisted in developing a clear mental model of the access system, possibly by the introduction of familiar and appropriate metaphors. Finally, the possibility of making errors should be minimized. Error correction should be simple and direct (Treviranus, 1994).

**Speech Recognition Systems**

Presently available speaker dependent speech recognition systems can be divided into two categories: static systems and adaptive systems (Smarte, 1989). When using static systems each word to be recognized must be explicitly trained. A static template of the trained word is saved and used to recognize the word. Some static systems (e.g., Voice Navigator™) allow vocabulary to be grouped in hierarchical groupings so that only a few words are compared at one time. In this way similar sounding words won’t cause confusion if they are in different groups. The user selects the group by speaking the group heading. Theoretically a user with only four discrete vocalizations can access 16 vocabulary items when the words are divided into groups of four. Utterances to be recognized need bear no resemblance to the intended word. The primary criteria for successful use are that the utterances be discrete from one another, and that they be repeatable and consistent. Thus a user who cannot produce any intelligible speech but can
make discreet, repeatable utterances can use a static speech recognition system to enter a limited set of commands or words.

Adaptive systems (e.g., Voicetype™, Dragon Dictate™, Kurzweil™) are trained through use and are frequently able to recognize or guess words that have not been spoken before by the user. These systems begin with a voice model (of normative English) which adapts to the individual user as the user dictates. Adaptive systems require that the user train a maintenance vocabulary which is used to control system functions, spell words which are not recognized and access non-text commands (approximately 200 vocabulary items). All other words are learned through use or are recognized using the adjusted normative models. Because the system is constantly learning and adapting, the user must be vigilant to errors. If errors are not corrected or corrected in the wrong way the voice model will be corrupted.

Adaptive systems are generally not suitable for individuals who do not use speech as a mode of communication (Treviranus, 1992). These users are unable to train the maintenance vocabulary of 200 words which are used to spell out words not recognized, correct misrecognitions and choose from a suggestion list. They are unable to use the voice recognition system for simple dictation as their vocabulary of discrete utterances is severely restricted. Also the effort required to adjust existing word models to speech which varies greatly from the norm makes the use of these systems impractical.

**Combining Speech Recognition with Scanning**

Combining limited vocabulary speech recognition with scanning to increase the efficiency of written communication for users with limited communicative acts has not been previously explored. The goal of this project was to investigate whether combining speech recognition with scanning would increase the computer input rate for children and young adults who are severely physically disabled and functionally nonspeaking and who use scanning as a means of controlling computers, and to determine whether assessments of the consistency and discreteness of speech as performed by speech language pathologists would serve as good predictors of the success of combining speech recognition with scanning.
Chapter 3:
Method

Experimental Condition

This study addressed the following questions:
1) To what degree will the addition of speech recognition affect input rate and accuracy?
2) How do physically disabled, nonspeaking children and young adults rank the ease of using scanning alone, and speech combined with scanning? Which input method do they prefer?
3) How do speech intelligibility, severity and type of dysarthria, as well as consistency and distinctness of utterances affect input rate and accuracy across the two access techniques?
4) What is the rate of improvement across the series of learning sessions?

The above questions were addressed by studying two conditions, each employing a different access technique. For each participant, both conditions used the same visual matrix, vocabulary, and feedback (figure 1). Matrices were designed for each participant and were based on the participant’s personal writing system.

Michelangelo fights Shredder. The turtles win. The turtles have pizza.

<table>
<thead>
<tr>
<th>Commands</th>
<th>Go away space period</th>
<th>? return read print</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page</td>
<td>Page1</td>
<td>Page2</td>
</tr>
<tr>
<td>Names</td>
<td>I April Donatello Leonardo Michelangelo Raphael Shredder</td>
<td></td>
</tr>
<tr>
<td>Actions</td>
<td>can go fight have know like love stab slash talk tell win</td>
<td></td>
</tr>
<tr>
<td>Little</td>
<td>a all and at but for if in is it not on that the to with</td>
<td></td>
</tr>
<tr>
<td>Describe</td>
<td>angry bad best great good fast hard loud many no strong</td>
<td></td>
</tr>
<tr>
<td>Things</td>
<td>car foot soldiers pizza shell turtles weapons</td>
<td></td>
</tr>
<tr>
<td>Places</td>
<td>apartment home office sewers street</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Sample personalized visual keyboard
The matrices were developed in consultation with an augmentative communication consultant and the participant's clinical team. Participants and their clinical teams chose vocabulary items they deemed necessary or desirable for written communication. The chosen words, phrases or letters were placed into logical categories using classifiers with which the participant was familiar. Classifiers or group headings for each row appeared in the left-hand column of the matrix. Words corresponding to group headings filled the rest of each row. Since the system was dependent upon grouping the vocabulary in meaningful classifiers, the participant's ability to use classifiers was assessed (using sorting tasks). Classifiers chosen by the participant, which the participant was able to comprehend were used. Participants had access to additional vocabulary by selecting page headings displayed in the top row. Selected words were sent to a text window above the matrix. The two conditions were as follows:

A) Scanning
The subject moved through and selected from the matrix using his or her accustomed mode of scanning and personal switch.

B) Scanning combined with speech recognition
In this condition utterances could be used to directly select a row within a scanning matrix or a specific item. With two participants an utterance was also used to replace the switch used in condition A. Participants selected the page, group or item by speaking an associated utterance or word approximation. If there was no distinguishable utterance for the desired page, group, or item the participant used a predetermined utterance or switch closure to command the system to scan the rows or items and used another predetermined utterance or switch closure to indicate that the desired row or item had been reached. To choose any item the participant first chose the group it belonged to and then the item. This reduced possible confusions among words recognized by the speech recognition system. The participant could use either scanning or voice at any stage in the selection procedure. If the speech recognition system failed to recognize an utterance, the participant could resort to scanning.

The study employed a within-subject counterbalanced design with repeated measures. All subjects participated in both conditions. One of two possible orders of presentation was randomly assigned to each participant. An equal number of participants were assigned each order of presentation. The design was chosen because of the heterogeneity and size of the subject group and to reduce the time commitment required of each participant.
The dependent variables studied included:
1) the net rate of input measured in correct entries per minute (a summary measure of rate and accuracy).
2) post hoc analyses of input errors,
3) subjective ease of use rating by the participants, and
4) personal preference rating by the participants.

Participants
Ten participants were chosen from the caseloads of local augmentative communication clinics. Trials with two participants were not completed due to illness. Participants met the following criteria:
1) their means of written communication included traditional orthography as part of the representational set,
2) their means of computer input was scanning with discrete switches,
3) their speech was inadequate to meet their daily communication needs, and
4) they could make three or more repeatable, but not necessarily intelligible, utterances which could be distinguished by a speech recognition system.

All participants were diagnosed with cerebral palsy. The ages of the participants ranged from 5 years to 33 years of age (See table 1). Experience with scanning systems ranged from 6 months to eight years. The reading/writing levels of the 5 older participants ranged from grade 1 reading/writing to grade 3 reading and grade 2 writing (as estimated by the participant's teacher or literacy tutor). All participants used some form of manual augmentative and alternative communication (AAC) system with the majority using partner assisted scanning to indicate selections. Two of the participants had used electronic AAC devices for face-to-face communication in the past but had ceased to use them. Table 1 lists the average rate of input using existing writing systems for each participant in items per minute (i.e., words, punctuation, pictures, symbols).
<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Intelligibility Rating</th>
<th>Number of Words Recognized</th>
<th>Reading level (grade)</th>
<th>Writing level (grade)</th>
<th>Scanning Rate Using Writing System (ipm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1-2</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>3</td>
<td>14 no groups 62 in groups</td>
<td>2</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>2</td>
<td>4 no groups 7 in groups</td>
<td>1-2</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>4</td>
<td>16 no groups 63 in groups</td>
<td>3</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>2</td>
<td>6 no groups 19 in groups</td>
<td>2</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1</td>
<td>3 no groups 7 in groups</td>
<td>prereading prrewriting</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>prereading prrewriting</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>prereading prrewriting</td>
<td>0.40</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Participant Profile

**Speech Screening**

An assessment of the participants’ oral motor function, phonation, and general intelligibility of speech was performed by a qualified speech language pathologist using standardized measures or standardized measures appropriately modified to accommodate the severe dysarthrias of the participant group (refer to Table 1, Intelligibility Rating and Appendix 1 Speech Status). The speech intelligibility (at a single word level) of each participant was rated by a Speech Language Pathologist on a ten point scale where 1 was completely unintelligible, 5 was somewhat intelligible and 10 was very intelligible. The utterances of all participants were rated as largely unintelligible.

**Speech Recognition Screening**

Each participant was also tested for the number of utterances recognized by the speech recognition system from a predetermined list of 30 words (refer to Table 1, Number of Words Recognized). Recognition was tested with the words grouped into logical groupings and in one large list. When grouped, the task of recognizing a given word is simplified as the choice of possible words spoken is limited; however, the user must speak both the group heading and the intended word. None of the participants would be considered candidates for using standard speech recognition systems.
Speech Recognition System

Speech recognition was performed by the Voice Navigator™ recognition unit, a device specifically designed for the Macintosh. This system offered a completely user definable vocabulary. Vocabulary could be grouped in a hierarchical fashion thereby reducing the number of words compared at one time. The vocabulary and word models were static and did not adapt through use but were specifically entered and trained. Recognition confidence required for the system to guess a word, microphone gain, and the number of models for each word were adjustable.

A static speech recognition system was chosen over an adaptive system for several reasons. Large vocabulary, adaptive systems appeared to be unsuitable for the purposes of this study and the population being studied. When using adaptive systems the vocabulary and therefore the words which appear in the suggestion lists could not be restricted, this caused unwanted and frequently unfamiliar words to appear in the suggestion list. As the system would attempt to find a likely match from a very large list of possible words the probability of predicting the intended word was reduced. Accurate recognition of untrained words was very low as recognition was initially based on models of normative speech and as the words spoken by the participants frequently bore little resemblance to the word models. Adaptive systems were inappropriate for user's whose literacy was not well established as adjustments to the word models depended upon the ability to spell. The speech capabilities of the participants was inadequate to train the approximately 200 word vocabulary required to maintain the adaptive system models.

Testing Software

The scanning interface was developed on the Macintosh™ SE/30 computer using HyperCard™. Feedback for scanning and voice recognition entry was identical. The entire row and then the item were highlighted. Auditory scanning of the rows and items was a selectable option. An immediate tone occurred upon selection followed by speech echoing of the selected category or item name. Any row or item could be selected by either scanning or voice recognition. The system was designed to recognize utterances even when the cursor was scanning through the text. Selected items appeared in a text window. Generated text could be printed or read out using the MacIntalk voice synthesizer.
Multimedia recordings were used to enhance the text. These included both audio recordings and digitized graphics. For example, recordings of the Ninja Turtles and a graphic image of that Ninja Turtle would appear upon selection of one of the Ninja Turtles. Opening cards of the participant's stack were also personalized using multimedia recordings.

**Procedure**

Participants took part in two introductory sessions followed by eight learning sessions for each condition. A short copy type task (12 to 15 items) was included at the end of each learning session during which rate and accuracy data was collected. In order to minimize interference effects, the conditions were separated by at least one week during which no training was provided. All participants received an honorarium to help defray the costs of taking part in the research.

**Introductory Sessions**

During the introductory sessions, the participant was introduced to the scanning matrix, the classifiers, and the available functions. In condition B the speech recognition system was also trained during the introductory sessions.

Vocabulary to be trained was chosen from the participant's matrix and trained in order of priority. Selectable items and commands were prioritized according to their frequency of use. Thus priority was given in the following order:

1) command words such as "undo" and a command which took the cursor to the second half of the rows or items within a row,
2) row or page headings,
3) words, giving priority to words which required the highest number of scanning steps to reach.

During the training the participant was simply prompted to "say the word." Whatever utterance was produced was recorded as a model for the vocabulary item. Three models were recorded for each word. Following the training of each word, speech recognition was tested to determine whether there were confusions. If confusions occurred with previously trained words, the lower priority word was retrained. If confusions continued to occur, the lower priority word was eliminated from the speech recognition vocabulary (to be retrieved using scanning). Vocabulary to be trained was not chosen by taking an inventory of the participant's reliable utterances. This would cause the recognizable
spoken utterances to dictate the vocabulary available for writing. Participants were also not asked to assign repeatable discrete vocalizations to high priority items. If this strategy was used, participants would need to remember the associated utterance, thus adding an additional layer of complexity.

Learning Sessions
During the learning session each participant engaged in motivating, age-appropriate tasks which maintained their attention and which required a minimum of skills not associated with controlling the access technique. Two sets of equivalent tasks were chosen with the participant prior to the learning sessions and randomly assigned to the two conditions. Activities included: fan mail to New Kids on the Block, letters to a local baseball team, letters to the prime minister demanding his immediate resignation, complaint letters to the accessible transit service, as well as other writing tasks chosen by the participants.

Younger Participants
Trials with the three youngest participants differed from the other participants in several respects. As these children had never experienced copy type tasks and had difficulty grasping the necessity to copy text quickly and accurately, two game-like programs were designed. The first task was a pizza-making task (see Appendix 4). The scenario described to the participants was that they were pizza chefs in a popular pizza chain. They were to fill customer orders as quickly and accurately as possible or the customer would not pay. The child was to choose from two rows of pizza ingredients. The list of ingredients was displayed graphically, the child was also given spoken prompts. Once the item was selected using either scanning or voice it would appear on the pizza on the screen. Items were classified as “food” or “silly.” The three participants assisted in choosing the items to be included in the ingredient list. This task was used during 4 learning sessions for each condition.

The second task was a shopping task (an activity reported by parents and teachers as being highly motivating for all three participants). In this task the child chose a store to enter from a row of storefronts using either scanning or voice. The screen would then display two rows of items available in the chosen store (see Appendix 2). The three participants were asked to fill a shopping list as quickly and accurately as possible. Buying the wrong items would mean they ran out of money. Items "purchased" would appear in a shopping cart on the screen.
Decision Criteria

When using scanning and speech recognition the participants were required to choose which access technique to use to select each category and item. In pilot trials with two users it was found that participants performed better when they were given clear decision criteria. Given this finding clients were given clear directions regarding when to use speech and when to use scanning. These decision criteria depended on the number of trained vocalizations the participant had available. Participants with a large number of vocalizations were directed to try voice first and if it failed to resort to scanning. For participants with a moderate number of recognized vocalizations, the words or categories that could be selected using voice were displayed in bold. They were then directed to select bold items using voice and other items using scanning. Speech access was usually restricted to high priority commands and category headings for participants with less than 8 vocalizations.

To assist the younger participants in deciding whether to use voice or scanning to retrieve an item, the store analogy was elaborated. The participants were told that magic passwords were being entered when training the store names. They were instructed that in order to get into the store they had to say the magic password (magic passwords being their individual way of saying the name of the store). These participants were quite pleased when it was demonstrated that the researcher could not enter the store by saying the same word. They also had no difficulty in remembering which stores they did not have magic passwords for and must walk to (access through scanning).

Two participants developed their own strategies to exploit the benefits of both selection techniques. These participants would frequently hit the switch to begin the scanning process and then attempt to say the desired word. If their vocalization was not recognized the cursor was already on the way to the desired item.

Control for Order Effect

To control for order effect, one of two possible orders was randomly assigned to each participant (condition A then B or condition B then A). An equal number of participants took part in each order.
Measures

Copy Type Tasks

The copy type task completed at the end of each learning session consisted of one or two sentences from 1 to 16 in length (words and punctuation). The sentences were chosen by a clinician blind to the location of the words in the scanning array and the words the participant was able to access using speech. The clinician was aware of the available vocabulary, reading level and interests of the participant. The sentences related to the theme of the learning sessions. Posthoc analysis of the sentences chosen showed that they were balanced across conditions according to a scanning difficulty rating. This scanning difficulty rating was obtained by scoring all words within a copy type task according to the number of scanning steps required to reach each word and calculating a sum of scores for each copy type task.

The researcher recorded the rate of input in items (words/punctuation) per minute (see Appendix 3 for sample scoring sheet). The participant was asked, and if necessary, prompted to correct all errors. Errors and corrected errors were recorded. In condition B non-recognitions and mis-recognitions were recorded as was the method of selecting each item (scan or voice).

Error Classification

Errors were recorded and categorized as one of the following:
1. Scanning error: inaccuracy in timing switch activation or switch release.
2. Copy error or omission: inaccuracy in copying the text to be copied.
3. Selecting the wrong category: the wrong category was chosen for the intended word, or
4. Mis-recognition of utterance: the voice recognition system misrecognized the utterance.

Non recognitions of utterances were also recorded to calculate the average number of repetitions required prior to recognition. Non-recognitions were not included in the error calculation as they can be largely attributed to the system and as they did not result in the entry of unwanted words.

Subjective Ease of Use and Preference Ranking

Following completion of all trials for one condition subjects were asked to rate the access technique for ease of use. A 5 point, Likert-type scale was used: very easy, easy, neither
easy nor difficult, difficult. very difficult. Following completion of all trials subjects were asked which access technique they preferred.
Chapter 4:
Results

Order Effect
Data was analyzed using repeated measures analysis of variance (repeated Anovas) to determine whether there was an order effect. No significant order effect was found (F= 1.93, p= 0.2138, see Appendix 2).

Difference in Net Rate
Differences in net rate (correct selections per minute) between condition A and condition B were analyzed using paired t-tests (see Table 2). When all eight participant results were included in the analysis, a significant difference was found between the overall net rate of 1.34 items per minute in condition A and the overall net rate of 2.73 items per minute in condition B (T = 3.47, p= 0.01) (see Appendix 2). A greater difference was found when the three youngest participants were excluded from the calculations (T= 6.10, p= 0.0037) (see Appendix 3). Repeated measures analysis of variance with the order term excluded yielded identical results.

Difference in Accuracy
Analysis of the number of errors made in condition A versus condition B using a paired t-test found no significant difference in the overall error rate. Table 2 shows the total number of errors for the two conditions. It also shows the average number of repetitions required before a word was recognized by the speech recognition system and the average number of mis-recognitions or words recognized incorrectly per copy type task. As can be seen from the data, participants frequently needed to say a word more than once before it was recognized by the system (this may be partially due to fluctuating volume and duration of utterance which would cause the system to reject the utterance as not within the trained set of recognizable words). Average number of words misrecognized varied from 0.1 to 1.7. This reflects the discreteness of the utterances (i.e., if utterances are not sufficiently different from one another one utterance can be more easily confused for another).
Correlations Between Speech Intelligibility and Rate Difference Score

Correlation analysis was performed using Pearson Correlation Coefficients to determine whether there was a correlation between the rate difference scores (difference in participant's average rate in items per minute between condition A and B) and the intelligibility rating to determine whether speech intelligibility correlated with success in using scanning combined with voice recognition. Intelligibility rating and difference scores were found to be significantly correlated ($r = .81$, $p = .01$). The number of words recognized by the recognition system and the difference scores were not significantly correlated ($r = .65$, $p = .07$). This may be due to the small sample size.

Individual Participant Results

Figures 2 to 9 show the individual rates in items per minute for both conditions across the eight learning sessions for each participant. A typical learning curve was not evident with any of the participants. All participants input at a higher rate using condition B on the first learning session. The 3 younger participants (participants 7, 8, 9) showed more erratic results in subsequent learning sessions.
Figure 2: Participant 1, items per minute, scan and voice and scanning alone.

Figure 3: Participant 2, items per minute, scan and voice and scanning alone.

Figure 4: Participant 3, items per minute, scan and voice and scanning alone.
**Figure 5:** Participant 4, items per minute, scan and voice and scanning alone.

**Figure 6:** Participant 5, items per minute, scan and voice and scanning alone.
Figure 7: Participant 6, items per minute, scan and voice and scanning alone.

Figure 8: Participant 7, items per minute, scan and voice and scanning alone.

Figure 9: Participant 8, items per minute, scan and voice and scanning alone.
Subjective Ease of Use and Preference Ranking

Subjective ease of use rating and access technique preference are shown in Table 3. The majority of participants preferred to use scanning combined with voice. The ease of use rating for both conditions ranged from difficult to very easy. Generally, participants ranked the two conditions as being equally easy or difficult to use.

<table>
<thead>
<tr>
<th>Participant</th>
<th>A: Scanning Alone</th>
<th>B: Scan &amp; Voice</th>
<th>Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>very easy</td>
<td>easy</td>
<td>Scan and Voice</td>
</tr>
<tr>
<td>2</td>
<td>neither easy/difficult</td>
<td>very easy</td>
<td>Scan and Voice</td>
</tr>
<tr>
<td>3</td>
<td>easy</td>
<td>easy</td>
<td>neither</td>
</tr>
<tr>
<td>4</td>
<td>easy</td>
<td>easy</td>
<td>Scan and Voice</td>
</tr>
<tr>
<td>5</td>
<td>neither easy/difficult</td>
<td>neither easy/difficult</td>
<td>Scan and Voice</td>
</tr>
<tr>
<td>6</td>
<td>difficult</td>
<td>difficult</td>
<td>Scan and Voice</td>
</tr>
<tr>
<td>7</td>
<td>easy</td>
<td>easy</td>
<td>Scan and Voice</td>
</tr>
<tr>
<td>8</td>
<td>neither</td>
<td>neither</td>
<td>Scan and Voice</td>
</tr>
</tbody>
</table>

Table 3: Subjective ease of use and preference ranking.
Chapter 5: Discussion

Rate

It appears that gains can be made in items selected per minute using scanning combined with speech recognition even with a very limited set of discrete vocalizations. Whether these improvements are sufficient to outweigh the additional training (of the speech recognition system and user) and equipment needed can only be determined by the computer user and clinical team. The gains are evident from the first session onwards suggesting that improvements in efficiency counteract the learning demands of using two selection techniques.

Accuracy

Although adding an additional access technique also added additional types of errors, (mis-recognitions of words and recognition of involuntary vocalizations as words), there was no significant decrease in accuracy from condition A to B. All participants made scanning errors which included inadvertent activations of the switch, inadvertent double hits of the switch, as well as timing errors such as hitting the switch too late or too soon. When directly selecting using voice, the opportunities to make scanning errors were reduced and replaced by voice recognition errors.

Difference in Rate caused by “undo” and “second” function

The difference in rate gains between participants who had very few vocalization (i.e., participant 1 and 3) and participants who had a relatively large number of recognized utterances is not proportional. Direct access to commands such as “undo” which undoes the previous action and “second” which bypasses half of the scanning array appear to have a far greater effect than direct access to individual vocabulary items. This may imply that far simpler voice interfaces which directly select these commands could achieve similar gains when combined with scanning.
Learning Trends

A typical learning curve was not evident with any of the participants (refer to Appendix 1). In condition B this may be due to the fact that voice training occurred during the introductory sessions. Utterances produced by the participants evolved with practice (i.e., additional articulators were added, and extraneous sounds were eliminated). The further the trials were from the initial training, the less representative the recorded models were of the utterances to be recognized. Thus mis-recognitions and non-recognitions increased. This effect may have counteracted any learning effect when looking for trends in the learning curve. As all participants were familiar with scanning prior to engaging in the study, significant improvements in scanning performance would not be expected.

Performance of Younger Participants

The three youngest participants did not make significant gains in writing rate when using speech recognition. This could be attributed to several factors. It became obvious during the trials that accuracy and rate were not priorities for these participants. Other variables influenced the results. Errors tended to be “silly” or more motivating items chosen over serious items. Food items in a pizza task were frequently replaced with bugs, or extra insects were added to the ingredients. Likewise in the shopping task participants usually picked their favorite items while filling the shopping list or explored stores they preferred. This would result in more errors and a slower net rate of input.

Speech Recognition System

During the course of the study, the Voice Navigator Speech Recognition System was upgraded. Unfortunately the “improvements” to the system made it unsuitable for the users participating in the study. Improvements to the recording and sampling quality of the speech recognizer meant that variations in speech production across voice training models were more accurately detected leading to the rejection of the training sequence by the speech recognition software. The elimination of a manual gain control made the system unsuitable for participants with low volume speech.

Developers of the Speech Recognition System were persuaded to make modifications to the software to accommodate the participants. These changes included a gain control override, a sampling quality adjustment and a tolerance adjustment. The tolerance
adjustment controlled the tolerance to variations in formants when recording the three models for each word. This tolerance adjustment, however, did not adjust the tolerance to variations in word and vowel duration. The researcher resorted to programming three separate word vocabulary items for the same word with one model each, to overcome rejections due to variation in duration, thereby tricking the system into viewing each duration variation as a different word.

The developers did not allow adjustment of the word boundaries (pause required between words). This resulted in the interpretation of words pronounced with hesitations or breaks in voicing as more than one word. User control of pauses required to mark word boundaries would greatly improve recognition for these participants. Better filtering of involuntary oral sounds produced when breathing, preparing to say a word or associated with hitting a switch would also improve performance. The researcher trained these sounds as null vocabulary items when possible.

Pronunciation of words by participants evolved over the course of the trials. These pronunciations more closely resembled the normative pronunciations. Extra sounds were dropped (e.g., mpa to pa) or articulators added (e.g., bo to bob). These shifts occurred despite the fact that they resulted in non-recognitions and were therefore discouraged by the system. These improvements are likely due to the increased use of speech.

Recommendations

In order to better accommodate the needs of this population, a voice recognition system for impaired speech is required. This system would have the following specifications:
1) a small user definable vocabulary with hierarchical groupings where the tolerance to variation is proportional to the size of the vocabulary,
2) adjustable word boundaries,
3) adjustable tolerance to variations in duration,
4) adjustable gain control,
5) filtering of involuntary oral sounds, and
6) word models which adapt with use.

Ideally, this system should allow the entry of two different sets of models: models of the user's present speech and target models which are closer approximations of the normative
pronunciation of the words. Improvements in speech would therefore not be discouraged by the system.

**Future Research**

In order to construct efficient multimodal input systems, several design principles need to be further investigated. One of the predicted drawbacks of using two modes of access is the added cognitive load of switching between modes of access and deciding which mode of access to use with each selection (Forren and Mitchell, 1986). This may be partially alleviated by employing simple decision criteria for assigning selectable items to modes of access. This study assigned speech recognition as an optional selection method wherever possible (with assignment to high priority items first), and scanning as the backup selection method. In some cases items to be selected by voice recognition were marked. Possible alternative strategies to be explored include: assigning modes to different communicative functions, restricting certain vocabulary items or commands to a single mode or modifying the effect of one mode with another mode. The optimal assignment will likely be influenced by the user's range of recognizable utterances.

Another question to be addressed is: how influential is the integration of feedback and the type of feedback given on the user's ability to deal with multimodal input. Is the additional cognitive load of using two access techniques significantly lessened by integrating the display feedback for the two methods? The results of this study may be influenced by the integration of the display feedback for scanning and scan and voice.

Lastly, methods and orders of introducing multiple input methods should be explored. Should the user gain mastery in one method before the next method is introduced? Are there specific skills related to the successful employment of more than one access method?
Individuals with very limited speech cannot use speech recognition as their primary or sole method of access; however, speech recognition can be used to augment other less direct methods of access. This can have a significant effect on rate of input for individuals with very few controllable actions. When speech recognition is combined with scanning faster speeds of text entry can be achieved than when using scanning alone. Accuracy is not compromised despite the introduction of new error types. Assessments of the consistency and discreteness of speech as performed by speech language pathologists serve as good predictors of the success of combining speech recognition with scanning.

Speed and accuracy do not appear to be important priorities for very young computer users. Other measures of success need to be found for computer users aged 6 and under. These may include attention to task, and ease of learning the access technique.

Presently available speech recognition systems do not adequately accommodate users with limited, non-normative speech. Further advocacy and consultation is required to insure that these users have optimal access to speech recognition systems.

This research demonstrates that computer users with severe motor impairments can use multimodal access techniques when careful attention is given to sources of feedback and integration of the modes of access. Software and implementation strategies developed through this project can serve as models for future investigation of multimodal access systems.
Chapter 7: References


Hypercard is a registered trademark of Apple Computer Inc.

Dragon Dictate is a registered trademark of Dragon Systems Inc.

Voice Navigator is a registered trademark of Articulate Systems Inc.
Voice Type is a registered trademark of IBM Corporation.
Appendix 1: Individual Participant Profiles
Participant 1

Speech Status
Consistency and discreteness of sounds which Participant 1 did make (mostly a central nasal vowel) were maintained. He often tried to maintain the number of syllables in the word (e.g., uh uh). Overall there was little differentiation in sound production. He had very few vowels and consonants. He had good intensity, but took great effort to initiate a sound. He was able to produce the nasals (e.g., m,n), fricative (e.g., f,h), stop (e.g., p) and vowels "uh", "eh", au "ough", a "ah". Most of the vowels were nasalized. Syllable forms observed included (where v = vowel, C = consonant, N= nasal): V, VC, CV, CVC.

Sample Copy Type Tasks
I will fly to Portugal in a Concord jet. My family will come with me.
I have a Wild-runner truck. I want to race my truck and win.

Selections per Minute: Scan and Voice and Scanning Alone

![Graph showing selections per minute for Scan and Voice and Scanning Alone over 8 sessions.]
Participant 2

The consistency and discreteness of her sound production was good for those sounds that she had. Poor breath support, low intensity levels, high pitch, voiced speech throughout, nasality, characterized her voice. Intensity varies depending on her ability to produce the word. Initiation is fairly quick. Overall, her speech intelligibility was poor because of low intensity, weak intraoral air pressure, continuous voicing, high pitch, small consonant and vowel inventory, and consonant substitutions. Word stress seemed to be intact, but poor breath support made the stress pattern difficult to hear. Fatigue may be a factor contributing to her reduced speech intelligibility. Participant 2's consonant inventory included stops (e.g., p, b), fricatives (f, v, s), nasals (e.g., m, n), liquid (l), affricates (e.g., ts). Vowels produced are a as in "ah", ae as in "hat", er, aI as in "eye", oI as in "boy".

Syllable forms observed included:
V, CV, VC, CVC, CVCC, CVCV, CVCCVC.

Sample Copy Type Tasks

I love (New Kids on the Block). I would like to go to your concert.
I would love to go to a romantic restaurant with you and get to know you.

Selections per Minute Scan and Voice and Scanning Alone

![Graph showing selections per minute for Scan and Voice and Scanning Alone over 8 sessions.](image-url)
Participant 3

Consistency and discreteness of sounds which this participant made were somewhat variable. There was slow initiation of and prolonged movement of the articulators, but the intensity level was good. There was adequate oral pressure contact of the articulators for the stop b. Since the rate was so slow, vowels were sometimes distorted. Voicing was present throughout speech. Consonants were the nasal m, stop b, approximation of g, and fricative s. Vowels were a "ah", i "ee", ae "at", "uh", au "ouch". He frequently turned these vowels into diphthongs by moving into a central vowel at the end. Syllable forms observed included V, CV, CVC, CVCCV, CVCCVCVCVC.

Sample Copy Type Tasks

I think Brian Mulrooney doesn’t know what he is doing to Canada. (Read)
Your boys are too rough! We are the best hockey players.

Selections per Minute Scan and Voice and Scanning Alone
Participant 4

Participant 4’s production of discrete speech sound was very reliable. His speech was characterized by a high pitched voice, continuous voicing, reduced breath support causing a very low intensity, but did reach articulatory targets. There was some degree of control over voicing. He produced several good word approximations. He marked the number of syllables. Fatigue may be a factor in his speech unintelligibly over a period of time. The consonants are stops (e.g., k,g,p,b,t), fricatives (e.g., f,h) liquid (l). Vowels are i "ee", "uh", au, as "ouch", a "ah".

Syllable forms observed are V, VC, CV, CVC, VCCV, CVCV.

Sample Copy Type Tasks

My sister will have a baby this week. I will visit her soon. I sent a letter last July. You have not answered my letter.

Selections per Minute Scan and Voice and Scanning Alone

![Graph showing items entered per minute vs sessions. The graph compares Scan and Voice (solid line) and Scanning (dashed line). The y-axis represents items entered per minute ranging from 0 to 5.5, and the x-axis represents sessions from 1 to 8. The graph shows a decline in items entered per minute for Scan and Voice from session 1 to 5, followed by an increase from session 5 onwards. Scanning remains relatively stable throughout.]
Participant 5’s speech sound consistency was good. He had limited sound production, however, he did make discrete sounds. He had poor breath support, however, he did get enough intraoral pressure and articulator closure to produce stops. Initiation was fairly quick, voicing fluctuated. Fatigue is expected to be a factor causing poorer speech sound intelligibility. Consonants produced were nasals (m,n,ng), stops (t,d, b, p, k, g), fricatives (f,h). Vowels were a central  "uh", i "ee", u "oo", i "ee", a "ah", el "day", ae "hat".

Syllable forms observed were V, UC, CV, CVC, VCV. Of note is his need to start his voice going, producing a nasal consonant in the position of the stop which he wants to articulate, NCVC (eg. mpay).

Sample Copy Type Tasks
I want to tell Wendel Clarke how much I like the Toronto Maple Leafs. The Pittsburgh Penguins played a terrible game against the Habs last night.

Selections per Minute Scan and Voice and Scanning Alone
Participant 6

Participant 6’s speech sound consistency was good for those sounds he could produce. His discrete productions were small in number. Breath support and intensity level were low. Fatigue may be a factor in his productions. There was slight voiced/voiceless distinction indicating some control of vocal cord function. The consonants which he made were stops (eg. g), fricatives (z as in garage, sh), nasal (n,ng), glide (w). Vowels were the central “Uh”, and “eh”.

Syllable forms were V, VC, CV, CCV.

Sample Copy Type Tasks
Crust, Cheese, peppers, snail, pepperoni
(Toy) teddy, (pets) snail, (sports) saddle, (clothes) hat, (fruit) grapes

Selections per Minute Scan and Voice and Scanning Alone
**Participant 7**

Participant 7's sound production was fairly consistent for those sounds he could produce. Initiation was slow, and there was little differentiation in sound production. He attempted to mark two syllables. Fatigue, lack of speech sounds, poor breath support and low intensity contributed to his poor intelligibility. Consonants were nasal m. The vowel he produced was a nasal central "uh", and an approximation of a diphthong as in ouch.

Syllable forms are V, VC.

**Sample Copy Type Tasks**
Crust, Cheese, peppers, snail, pepperoni
(Toy) teddy, (pets) snail, (sports) saddle, (clothes) hat, (fruit) grapes

**Selections per Minute Scan and Voice and Scanning Alone**
Participant 8

Participant 8 was consistent in his speech production, however, there was limited differentiation in speech sound production. Initiation of movement was very slow. He had poor breath support, with fluctuating control, contributing to poor initiation of voicing. Most sounds were hyper nasal. It appeared that fatigue was a factor in unintelligibility.

Consonant was nasal (m). Vowels were central "uh", nasal and ae "at". Syllable forms observed were V, VN, N

Selections per Minute Scan and Voice and Scanning Alone

![Graph showing selections per minute for Scan and Voice and Scanning Alone across 8 sessions with values ranging from 0.1 to 0.7.]
Appendix 2: Statistical Analysis
Table 1: Repeated Measures Analysis of Variance for Order Effect

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>1</td>
<td>6.77157031</td>
<td>6.77157031</td>
<td>6.19</td>
<td>0.0473</td>
</tr>
<tr>
<td>ORDER</td>
<td>1</td>
<td>2.11408203</td>
<td>2.11408203</td>
<td>1.93</td>
<td>0.2138</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>6.56183594</td>
<td>1.09363932</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Paired T Test Summary, All Participants

| N Obs | Variable | T       | Prob>|T| |
|-------|----------|---------|------|
| 8     | SCANVOIC | 4.5899676 | 0.0025 |
|       | SCAN     | 5.2921559 | 0.0011 |
|       | DIFF     | 3.4734766 | 0.0104 |

Table 3: Paired T Test Summary, 5 Older Participants Only

| N Obs | Variable | T       | Prob>|T| |
|-------|----------|---------|------|
| 5     | SCANVOIC | 15.4825332 | 0.0001 |
|       | SCAN     | 11.1336886 | 0.0004 |
|       | DIFF     | 6.0960454 | 0.0037 |

Table 4: Paired T Test Summary, 3 Younger Participants Only

| N Obs | Variable | T       | Prob>|T| |
|-------|----------|---------|------|
| 3     | SCANVOIC | 6.2793092 | 0.0244 |
|       | SCAN     | 13.9421883 | 0.0051 |
|       | DIFF     | 2.5343249 | 0.1268 |
Appendix 3: Sample Scoring Sheet
Appendix 4: Hypercard Stacks Used with Younger Participants
<table>
<thead>
<tr>
<th>Pets</th>
<th>Start</th>
<th>Quit</th>
<th>Scan Rate</th>
<th>Wait</th>
<th>Auditory scan</th>
</tr>
</thead>
</table>

Guitar TURTLE

<table>
<thead>
<tr>
<th>SLIPPERY</th>
<th>TURTLE</th>
<th>FISH</th>
<th>SNAIL</th>
<th>TADPOLE</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>FURRY</th>
<th>BUNNY</th>
<th>MONKEY</th>
<th>ELEPHANT</th>
<th>BIRD</th>
</tr>
</thead>
</table>
Guitar TURTLE

FEET
SNEAKER
OOPS
BOOT

BODY
T-SHIRT
SWIMSUIT
HAT
Guitar TURTLE

Sports  Start  Quit  Scan Rate  □ Wait □ Auditory scan

BALLGAME  BASEBALL  BOWLING  SOCCER

RIDES  BOAT  BIKE  RIDING
Guitar TURTLE

Fruit | Start | Quit | Scan Rate | Wait | Auditory scan

JUICY
PEAR
APPLE
GRAPES

FUN
CONE
FRUIT FLY
TREEHOUSE
<table>
<thead>
<tr>
<th>Toy</th>
<th>Start</th>
<th>Quit</th>
<th>Scan Rate</th>
<th>Wait</th>
<th>Auditory scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guitar TURTLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIDE</td>
<td>FERRIS</td>
<td>ROCKER</td>
<td>TRUCK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLAY</td>
<td>TEDDY</td>
<td>BALLOONS</td>
<td>GUITAR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Food    crust    cheese    cucumbers    peppers    pepperoni    pineapple    tomatoes
Silly    spider    worm    snail    fly    car    ant    caterpillar    tadpole    bug