The Effects of Five Discrete Variables on Human Performance in a Telephone Information System

by

Michele Marie Cary

Thesis proposal submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Industrial and Systems Engineering

APPROVED:

Robert C. Williges, Chairman

Dennis L. Price

Beverly H. Williges

November, 1993

Blacksburg, Virginia

Investigating the Effects of Five Discrete Variables on Human Performance in a Telephone Information System

by

Michele Marie Cary

Committee Chairman: Dr. Robert C. Williges
Industrial Engineering and Operations Research

(ABSTRACT)

This study examined the effects of five dichotomous variables on human performance using a computer-based telephone information system. The five variables were: speech rate (120 or 240 words per minute), length of input time-out (two or ten seconds), feedback (available or not available), wallet guide - a graphical representation of the information (available or not available), and the database structure (8x2 or 2x6). The research methodology implemented a one-half fraction of a 2^5 factorial design, requiring only 16 of the 32 possible treatment combinations.

Two tasks were included in this study: a search task and a transcription task. The search task consisted of each subject accessing an information system through a touch-tone telephone. The subject listened to the computer as it used synthesized speech to list available menu options. The search task continued until the subject found the target item. The transcription task consisted of listening to and typing an information message for each target item. The experiment ended when 16 target items were found.

Four dependent measures were used to evaluate user performance. The search task was evaluated with three measures: user added time (the amount of additional time the subject required to complete the search in excess of the minimum search time imposed by the system design); invalid key presses (the number of times undefined keys on the touchtone telephone were pressed during the search); and user added key presses (the number of additional, valid key presses the subject required to complete the search in excess of the minimum number of key presses required to complete the search). Only one measure was used to evaluate user performance of the transcription task:

transcription accuracy score (the number of words that each subject transcribed correctly).

The results show four of the five variables (speech rate, database structure, input time-out, and wallet guide) to have a significant effect on human performance. The following interactions were found to have a significant effect on at least one of the dependent measures: database structure by input time-out, database structure by wallet guide, input time-out by wallet guide, and speech rate by wallet guide. Twelve subjective ratings were also analyzed. The results show at least one of the 12 subjective ratings was significantly affected by speech rate, input time-out, or the database structure.

Perhaps the most important finding of this research is that complicated auditory information structures can be accessed easily if a wallet guide is provided. In addition to decreasing search time, a wallet guide reduces the number of search errors users make.

ACKNOWLEDGMENTS

I would like to thank the National Science Foundation for the support they provided throughout the duration of this project. I would also like to express appreciation to all of my committee members: Dr. Dennis Price, Ms. Beverly H. Williges, and Dr. Robert C. Williges for their patience, guidance, and insightful suggestions which enabled me to complete this research successfully.

A special thanks is due to my colleagues in the Human Computer Interaction laboratory: Joe Chase, Jimmy Wu, and Sung Han.

A special debt of gratitude is also due to all my friends and family members who kept harassing me until the finishing touches were complete. Thank you.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

Telephone-Based Information Systems

Computer-based information systems using synthetic speech are becoming as commonplace as the telephones through which they are accessed. Allowing consumers to access virtually any database of information, with or without the aid of a human has become a strategy for many businesses. In addition to retrieving telephone numbers, it is currently possible to access such complex and dynamic information as personal checking or savings accounts, credit cards, I.R.S. refund amounts, electronic mail, and daily weather forecasts in any major U.S. city. Virtually any source of information can be retrieved via the telephone. Unfortunately, technological availability has preceded technological research, causing many product designs to neglect user or system requirements. User acceptance has been high for synthetic speech used in electronic mail, information systems, and information retrieval (Gould and Boies, 1984; Schmandt, 1985). However, as the technology is continually implemented on such a wide scale without properly researching user or system requirements, many people complain of unnecessary or poorly organized messages (Elmer-Dewitt, 1989).

Often, telephone information systems with auditory speech displays utilize little research in the developmental phase of production (Guide and Boies, 1984; Kidd, 1982; Podgorny, 1985; Schmandt, 1985; Witten and Madams, 1977). Witten and Madams (1977) developed a menu-driven telephone inquiry service with an auditory display that allowed interactive information storage and retrieval. In 1984, Gould and Boise investigated a voice-mail system which allowed users to receive, edit, or send messages to other users. Another voice-mail system was developed by Schmandt (1985). This prototype used a two-key combination of the 12 keys on a standard telephone keypad to define the letters of the alphabet.

Even though little research has been done to define the proper design of a telephone-base information system, some guidelines do exist. Halstead-Nussloch (1989) described two categories of decisions that an interface designer must make: mode of communication and dialogue structure.

Mode of communication refers to the type of input and output available. User input can be accomplished by voice or by depressing keys on a telephone keypad. System output is either digitized or synthetic speech.

Dialogue style consists of two types: command based or prompted. Command-based dialogue occurs when the user generates commands without prompts. This mode of dialogue is best suited for frequent or expert users. Prompted dialogue occurs when the user responds to prompts delivered by the system and is usually applied to a system with novice users.

According to Halstead-Nussloch (1989), the type of telephone information system used for this experiment is categorized as having prompted dialogue, keypad input, and rule-based synthetic speech as output. For example, input was provided when users pressed the "*" key on the telephone keypad to show that they wanted to select a menu option. Output was given in the form of rule-based synthetic speech, which also served as a prompt that let users know what the next menu option was.

Sequential Research Paradigms

Human factors research traditionally employs experimental designs investigating only two, three, or at most four independent variables of interest. Although experiments investigating less than four variables permit the use of full-factorial designs and allow the researcher to investigate all possible interactions, this methodology has several limitations. First, it provides limited domain information by excluding many system imposed effects. Second, the results of small factorial studies lack the ability to generalize across many factors. Finally, this methodology often produces results that fail to explain a considerable amount of variation in the data. For example, after Simon (1973) reviewed over 200 studies published in Human Factors, he found that on the average, experiments involving less than four independent variables were unable to account for a large portion of data variability.

When trying to assess the effects of five or more variables, full-factorial designs become very inefficient . The amount of time, money, and effort required to collect all possible treatment conditions would overwhelm the most experienced researcher. There are several alternative designs to the full-factorial when trying to investigate the effects of many variables. For example, fractional factorials, Placket-Burmans, and central composites are experimental designs which minimize the number of treatment combinations (Myers, 1976). These designs can be used independently as a screening study, or they can be combined in a series of experiments for sequential research.

Screening studies are often fractional-factorial designs used to investigate the effects of many independent variables. Alpha is usually set at a higher level to avoid making a Type I error (i.e., rejecting an independent variable when it does have a significant effect). For example, Simon and Roscoe (1984) evaluated six independent variables when they investigated a transfer of training effect. Tatro and Roscoe (1986) investigated the effects of eight display factors on pilot performance. Both studies utilized fractional-factorial designs, required far less data points than full-factorial designs, and were therefore very economical.

Sequential research methodology is a carefully constructed series of experiments which allow the researcher to investigate multi-factor design issues. Williges and Williges (1989) present an integrated research paradigm which involves three stages of data collection: selecting, describing, and optimizing variables of interest. Each stage of the paradigm utilizes experimental designs which fulfill the requirements of that stage. The first stage consists of selecting the independent variables, and utilizes screening studies of Resolution IV (design Resolution is determined by the smallest interaction in the identity relationship). Stage two describes the variables of interest using spherical or cuboidal designs such as the central composite or fractional factorial. The third stage of the paradigm uses specific design points and canonical analyses to determine the optimal levels for the independent variables.

This study is part of the second stage of the Williges and Williges (1989) sequential research paradigm. The first stage of the research (variable selection) was conducted by Merkle and Williges (1988) and by Beaudet and Williges (1988). Merkle and Williges performed a validation study that determined 19 of the 100 initial variables of interest warranted further investigation. Beaudet and Williges performed a complicated screening study that analyzed 16 variables with only 32 design points by using a Hadamard design matrix (Diamond, 1981). Since the Hadamard design is Resolution IV, only main effects were investigated; two way interactions are confounded with each other. The results of the screening study found seven variables that had a significant effect on user performance.

During the second stage of the research paradigm, two studies described the relationships among these seven independent variables and their subsequent two-way interactions with respect to user performance. Four of these variables were continuous and were analyzed using a central-composite design to study main effects, interactions, and to recommend optimal values (Wu, 1989). Three of the seven variables had dichotomous levels (e.g. "Available" or "Not Available") and were combined with the two design parameters to analyze main effects and two-factor interactions in a one-half fraction of a 2^5 factorial design.

This study exemplifies the second stage of the research paradigm, describing relationships among the independent variables, and leads into the third stage of optimization. The problem domain chosen by Williges and Williges (1989) to apply the integrated research paradigm is to evaluate systematically the design of a telephone-based information system.

Independent Variables

Following the screening study by Beaudet and Williges (1989), this study consisted of five independent variables: speech rate, length of input time-out, database structure, wallet guide, and feedback. All of the settings for the independent variables were set at the same values specified by Beaudet

(1988) except for input time-out and speech rate. In order to investigate the design space more thoroughly, input time-out was adjusted to a low value of 2.0 seconds and a high value of 10.0 seconds; speech rate had high and low values of 240 and 120 wpm, respectively.

Speech rate. Research on the proper speech rate to use has yielded conflicting results. Waterworth and Lo (1984) evaluated six different speech rates from 63 wpm to 150 wpm and found no significant difference in intelligibility. These speech rates were at low levels and no conclusions can be drawn concerning the intelligibility of speech rates higher than 150 wpm.

Simpson and Marchionda-Frost (1984) studied the effects of speech rate in a video game used to simulate a flight task. They used subjective ratings and found most users preferred a speech rate of 156 words per minute. Merva (1987) found subjects transcribed sentences more accurately at a speech rate of 180 wpm rather than 150 wpm or 240 wpm. However, subjects transcribing speech at 240 wpm showed such an improved performance over time that by the end of the experiment, they were approaching performance levels of the 180 wpm group.

Other studies validating the Merva (1987) finding that subjects transcribed sentences more accurately at 180 wpm were performed by Beaudet and Williges (1988) and Herlong (1988). Beaudet (1988) found synthesized speech at 180 wpm produced fewer transcription errors than did speech at 240 wpm. Herlong (1988) also conducted an experiment investigating speech rates of 180 wpm and 240 wpm with similar results.

Input time-out. The amount of time available for the subject to respond (i.e., invoke a key press) before the system presents another menu item defines input time-out. Many studies set input time-out without regard to its effect on user or system performance (Gould and Boies, 1984; Kidd, 1982; Schmandt, 1985). Beaudet (1988) investigated input time-out at two and four seconds, while Wu (1989) manipulated input time-out at two, three, six, nine, and ten seconds. The results of both studies show the logical effect that an increase in input time-out will increase absolute search time. However, the interaction of a long or short input time-out coupled with a simple or complex database structure is unknown.

Database structure. The structure of the database describes the hierarchy of menu levels used to access the information system. Menu-based computer systems are a popular means of human-computer interaction.

A major difference between an auditory and a visual display is that the information in an auditory display is presented serially, or one item at a time (Halstead-Nussloch, 1989). This limitation often requires the information to be organized in a hierarchical (tree) structure. The question then arises as to what type of hierarchy is best suited for an auditory display.

Designing the optimal database structure involves solving a trade-off issue between the depth (number of menu choices) of the hierarchy and the breadth (number of decisions) of the hierarchy. Menu depth and breadth are inversely proportional; as menu depth increases, menu breadth decreases. Several studies with visual displays have examined the depth/breadth tradeoff.

Miller (1981) tested 64 options with breadth values of 2, 4, 8, and 64, and found a "U" shaped function with respect to search time. These results suggest that the optimal breadth value falls between two and 64, but give no indication of where. Using the same database as Miller, Snowberry (1983) found a shallow depth with many menu options provides the best result in terms of speed and accuracy.

Kiger (1984) investigated auditory database structures and the depth versus breadth trade-off with an 8x2 and a 2x6 database structure. The results showed the 8x2 database to be superior to the 2x6 structure in terms of speed and accuracy. These two opposing hierarchical structures were also investigated by Beaudet (1988), which yielded similar results.

The database structures were uniquely designed so that they differ in depth and breadth but contain the same number of end (target) nodes. See Figures 1 and 2. Although the database structures have virtually the same 64 target items and corresponding information messages, main menu items and key words are different for each database. For example, there are only two main menu items in the 2x6 hierarchy ("Household" and "Fashion"), while the 8x2 database has eight main menu items. A key word is a title for a group of related store items (e.g., "fashion" is the key word for "clothing" and "accessories"). The 8x2 hierarchy contains two- or three-word clauses to represent each key word (e.g., "Men's clothing"). The 2x6 hierarchy, however, has many more menu levels and was able to breakdown the two- or three-word clauses into separate menu items, represented by a single word (i.e., "Clothing" is a key word for "Men's").

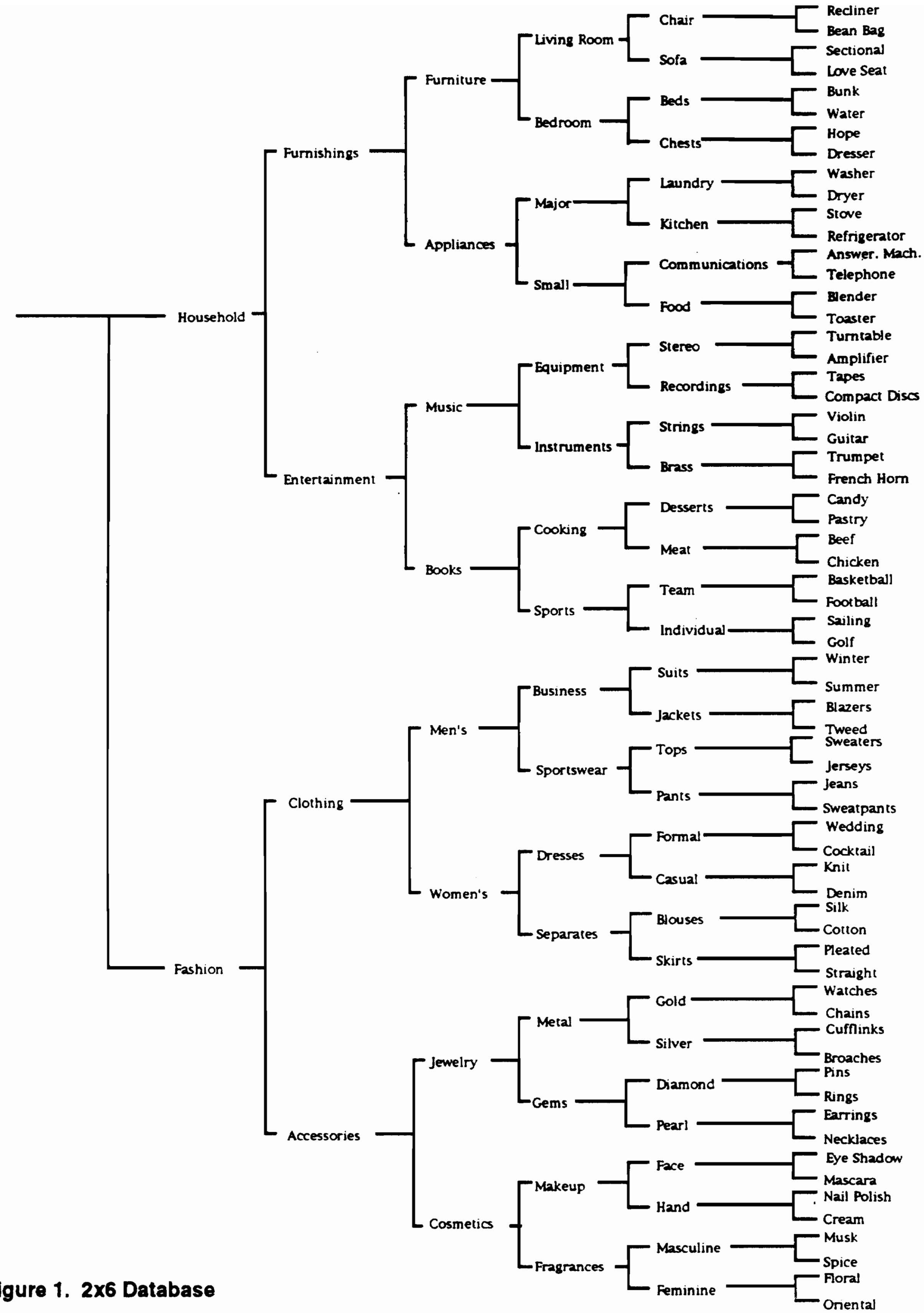


Figure 1. 2x6 Database

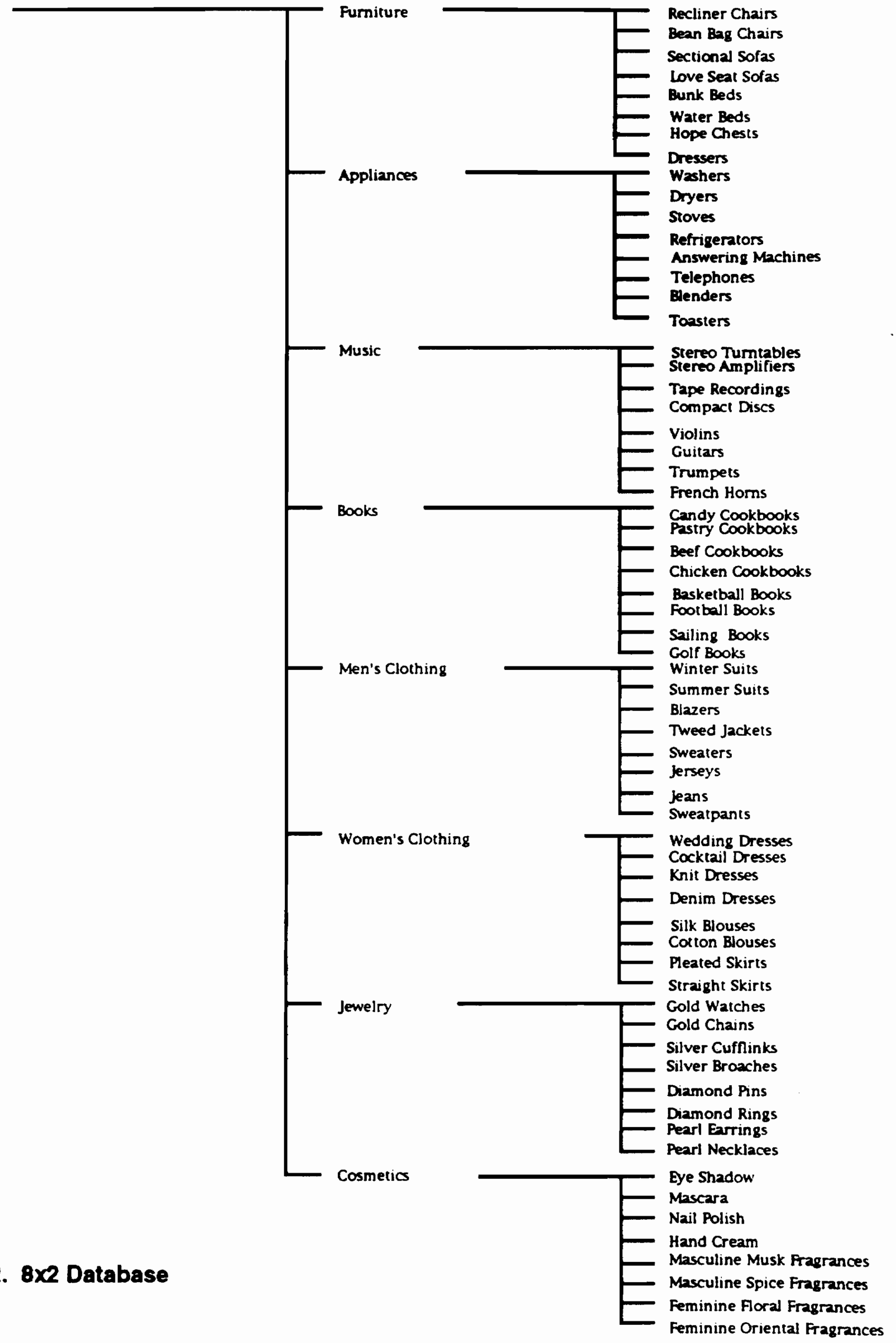


Figure 2. 8x2 Database

Wallet Guide. The wallet guide is a graphical representation of the database structure to which the subject is exposed. For example, if the subject had to search the 2x6 database, then a copy of Figure 1 (enlarged 100 %) was placed beside the telephone for them to use. A wallet guide is similar to user aids such as keypad overlays or command templates used in previous research (Podgorny, 1985; Witten and Madams, 1977).

Offering a visual representation of the database, such as a wallet guide, may help the user visualize the database structure. Additionally, a wallet guide should help prevent novice users of a hierarchical menu structure from making mistakes in their initial searches (Billingsly, 1982). Beaudet (1988) used a wallet guide to investigate the value of a visual representation of the database system to help subjects locate target items. His research found the use of a wallet guide decreases search time.

Feedback. Other than the tone of a push-button telephone, the design of most phones provides no feedback for the user when keys are depressed. The tone does not supply the user with specific information regarding what action was just selected. Spoken feedback provides this information and has been recommended as a system attribute for telephone-based information systems (Halstead-Nussloch, 1989).

Purpose

As part of a sequential research paradigm, the purpose of this study is to expand upon the screening study performed by Beaudet (1989), and to contribute to the final stage of the research paradigm: optimizing the levels of the independent variables. To expand on the screening study, a Resolution V design was used to isolate main effects and two-way interactions; three-way and higher order interactions were assumed to be negligible. The levels of the independent variables were set at values which would aid data collection of the optimization stage.

METHOD

Experimental Design

This study examined the effects of five independent variables on human performance in the design of a telephone-based information system. The five variables are: speech rate, input time-out, database structure, wallet guide availability, and feedback. Table 1 lists each variable and its corresponding levels. Since each variable had two levels, a 2^5 factorial design was used.

Only main effects and two-way interactions were investigated in this study, therefore, a 1/2 fractional factorial was used to reduce the number of treatment combinations from 32 to 16. A Resolution V design was constructed by sacrificing the five way interaction as the defining relationship. Table 2 shows the defining relationship, the two fractional factorial designs, and the aliasing which takes place in such a design. (For a more complete explanation of constructing a 1/2 fraction of a 2^k design, see Myers, 1976.) The 16 treatment combinations used to evaluate the five main effects and ten two-way interactions are shown in Table 3. Without any replication this design is fully saturated, meaning the degrees of freedom are equal to the number of data points. Therefore, replicating the design was necessary to investigate all 15 effects. A power analysis was conducted to determine the number of replications required (Keppel, 1982). With three replications per cell, beta was found to be approximately 0.12 when alpha equals 0.05. Appendix I shows the power analysis calculations. Finally, to avoid any effects due to training, and because age and gender are ongoing issues of the sequential research paradigm, a between-subject design was used.

Table 1

Levels of Discrete Variables

Independent Variables:	Low Level	High Level
Speech Rate (SR)	120 wpm	240 wpm
Input Time-out (IT)	2 sec.	10 sec.
Database (DB)	2 x 6	8 x 2
Wallet Guide (WG)	Not Available	Available
Feedback (FB)	Not Available	Available

Table 2

Identity Relationship and Ailiasing Structures

I = SR x IT x DB x WG x FB

Main Effects	Aliases
SR	IT x DB x WG x FB
IT	SR x DB x WG x FB
DB	SR x IT x WG x FB
WG	SR x IT x DB x WG
FB	SR x IT x DB x WG

Two-Factor Interactions	Aliases
SR x IT	DB x WG x FB
SR x DB	IT x WG x FB
SR x WG	IT x DB x FB
SR x FB	IT x DB x WG
IT x DB	SR x WG x FB
IT x WG	SR x DB x FB
IT x FB	SR x DB x WG
DB xWG	SR x IT x FB
DB x FB	SR x IT x WG
WG x FB	SR x IT x DB

Table 3

Treatment Combinations

Treatment Combinations	Independent Variables SR	IT	DB	WG	FB
1	240	2	8x2	NA	NA*
2	120	10	8x2	NA	NA
3	120	2	2x6	NA	NA
4	120	2	8x2	A	NA
5	120	2	8x2	NA	A
6	240	10	2x6	NA	NA
7	240	10	8X2	A	NA
8	240	10	8X2	NA	A
9	240	2	2X6	A	NA
10	240	2	2X6	NA	A
11	240	2	8X2	A	A
12	120	10	2x6	A	NA
13	120	10	2x6	NA	A
14	120	10	8X2	A	A
15	120	2	2X6	A	A
16	240	10	2x6	A	A

* NA - Not Available
A - Available

Method

Subjects

This study involved the use of 48 native, English speaking students at Virginia Polytechnic Institute and State University. To select subjects with minimal exposure to synthesized speech, only those who had not participated in a similar study, and were not regularly exposed to synthesized speech were selected. Participants consisted of twenty four male and twenty four female students, with an average age of 21.4 years and 19.8 years, respectively. As an incentive to particpate in the study, subjects were paid five dollars per hour.

Experimental Apparatus

A VAX 11/750 Mainframe system was used to present the tasks and record the data in real-time. Two DEC VT220 terminals were used throughout the experiment. The subject's station consisted of one VT220 terminal that provided target information and cued the subject when the search task was ready to begin. The experimenter's station also consisted of a VT220 terminal to initialize each session and to monitor data input.

The speech synthesizer used in this study was a Digital Equipment Corporation DECtalk, version 2.0. DECtalk was selected based on its superior performance when compared to three similar speech synthesizers (Green, Manous, Pisoni, 1984). The pronunciation by DECtalk was not enhanced by the use of manual phoneme or stress adjustment. However, compound words such as "sweat-pants" and "foot-ball" were entered with hyphens to reduce mispronunciation. The voice of Perfect Paul was used for the entire experiment; all other parameters were set at the default values.

The telephone system used in this study was chosen because it contains all of the necessary features (i.e., speaker, automatic re-dial) and was used during previous experiments of the research paradigm. A Panasonic VA-8205 Easaphone speaker telephone, with a volume control set at 64 dB(A) was used. Testing occurred in a quiet room with background noise of approximately 40 dB(A).

A JVC GX-S700U color video camera and a Sony Trinitron color monitor were used to monitor the subjects during the experiment. A General Electric video cassette recorder (model VG-7520) and television set were used to present subjects with a detailed explanation and demonstration of the task. In addition to the video instructions, all participants received a list of instructions, (shown in Appendix II).

Procedure

A summary of the experimental procedure is given in Table 4. It is divided into four sections: the welcome and orientation; the instructions and practice; the experimental task; and the post experimental session. Each section is described in detail below.

Welcome and orientation. Each subject received a brief description of the experiment and was given an informed consent document to sign (Appendix III). Those subjects who agreed to continue then completed a questionnaire (Appendix IV). Next, participants underwent a hearing test to determine if hearing in both ears was sufficient. A Belltone 109 audiometer was used to administer the hearing test. The hearing test consisted of three, one second pulsed tones at 26 dB(A) for 250, 1000, 2000, and 4000 Hz. Each frequency was presented to the subject's right or left ear; subjects were required to indicate they heard each tone by raising their hand. To pass the test, subjects must have indicated they heard two of the three pulsed tones. All subjects passed the hearing test and continued with the experiment.

Instructions and practice. Following the hearing test, an information system for a fictitious department store was called using the speaker telephone and the automatic dial feature. DECtalk's Perfect Paul gave a brief introduction (e.g., "Hello, and welcome to Hokie Wholesale") and read 17 instructions to the subject as they followed along with a written guide. After the subjects listened to and read the instructions, the experimenter watched a pre-recorded video tape with them. The video provided further instructions and a demonstration of the search task. When the video finished, the experimenter answered questions regarding the procedure of the experiment and asked the subject to

Table 4

Experimental Procedure

Welcome and Orientation	(~15 minutes)
Informed Consent	
Subject Demographics	
Hearing Test	
Instructions and Practice	(~20 minutes)
Introduction	
Video Instructions	
Written Instructions	
Subject Explains Task	
Practice Targets	
Experimental Task	(~45 minutes)
Targets 1-8	
Search Task	
Transcription Task	
Target Ratings	
Rest Period	
Targets 9-16	
Search Task	
Transcription Task	
Target Ratings	
Post-experimental Ratings	
Post Experimental Session	(~15 minutes)
Debriefing	
Payment and Dismissal	

describe the task. The subject then depressed the space bar causing the DECtalk to speak the keypad functions available to the subject. All instructions given by DECtalk during this period were spoken at the same rate as the treatment condition specified.

The subject initiated two practice searches by pressing a key on the speaker phone to call the department store database. Perfect Paul's voice again greeted and welcomed the subject to Hokie Wholesale and offered a review of the keypad functions. The first practice target appeared on the subject's terminal for 15 seconds, after which time a "Ready" prompt was displayed. The subject was cued to "Begin searching," while Perfect Paul spoke the first menu item.

When subjects heard a key word related to the target item, they selected the key word by pressing the "#" sign on the telephone keypad. This input caused the system to move to the next level of menu items. Perfect Paul responded by speaking the next (lower) level of menu items related to the selected key word. The search continued in this fashion until the target item was found. However, if during the course of the search, subjects wished to return to the main menu level, to back up one menu level, or to select the HELP menu to hear the key functions, they could press the "0" key, the "*" key, or the "#" key, respectively. If subjects arrived at a store item which was not the target, Perfect Paul said, "At store item _______, continue searching." Subjects pressed the "#" or the "*" key to continue the search.

Once the correct target item was selected, subjects were prompted to press the "2" key to hear the information message. All information messages had a ***subject verb preposition object*** sentence structure and were divided into one of four categories: location, price, availability, or information (Merva, 1987). see Table 5. Transcription accuracy was scored on the first and last two words of each sentence. The information messages were selected from the results of a pre-test based on the following criterion: at least one word in an information message was missed by at least one subject, but no words in the information message were missed by all subjects (Beaudet, 1988). After the computer spoke the information message associated with the store item, the subject transcribed the message by typing it onto the computer terminal.

After transcribing the message, the subject responded to a series of questions appearing on the screen. The first question asked the subject to determine subjectively the certainty of their transcription from 1 (very certain) to 7 (very uncertain). In the second question subjects determined the difficulty in understanding the information message from 1 (very certain) to 7 (very uncertain). The last question asked subjects to determine the difficulty in locating the store item from 1 (very difficult) to 7 (very easy). Examples of the subjective rating scales are found in Appendix V.

When subjects completed the last subjective rating, the second practice target appeared on the screen. The procedure continued exactly as before. Upon completion of the second target search, the system hung up. At this time the subject had a break and asked questions regarding proper task procedure.

Experimental task. After the second practice target, the experimental session began. The subject again dialed the number of the information system and heard Perfect Paul greet and welcome him/her to Hokie Wholesale. Searches proceeded in the same manner as the practice sessions. Each information message was followed by three subjective ratings. A

Table 5

Sentence Structure

LOCATION:	Subject(s)	*is(are)*	*in* *on* *near*	object.
PRICE:	Subject(s)	*is(are) reduced*	*by* *for*	object.
AVAILABILITY:	Subject(s)	*is(are) available*	*at* *by* *in* *with*	object.
INFORMATION:	Subject(s) Subject(s)	*is(are) allowed* *is(are) required*	*on* *at* *by* *in* *on* *to* *for* *from* *within*	object. object.

mandatory rest period was scheduled after the last subjective rating of the eighth target search. "Take a Break" appeared on the screen for one minute before the subject was prompted to continue searching for the next store item. Subjects continued in this manner until a total of 16 target items were found and the corresponding information messages were transcribed. Appendix VI lists all target items and information messages.

Post experimental session. When the subjects completed all 16 targets, they were prompted by the computer screen to rate the telephone information system on eight additional features. Each preference measure used a seven point bi-polar adjective rating scale. The preference measures are listed below.

- o ease of use of the information system
- o intelligibility of the computer voice
- o naturalness of the computer voice
- o speech rate of the computer voice
- o available time for user inputs
- o complexity of the database system
- o worth of the wallet guide *
- o worth of feedback *

* (only when treatment condition was available)

When all post-experimental ratings were answered, a debriefing session occurred. The experimenter asked the subjects several questions regarding their opinion of the information system, their impression of the database structures, and of the synthetic speech. All subjects were given an explanation of the experiment, thanked for their participation, and paid for their time.

Performance Measures

The experimental procedure involved two tasks: the search task and the transcription task. The search task involved two groups of performance measures: search time and search accuracy. The transcription task was evaluated by the strict scoring method outlined in Merva (1987), Beaudet (1988), Herlong (1988), and Merkle (1988). The performance measures for the search task are described below in greater detail.

Search time. Total search time is the amount of time necessary to locate the target item once the prompt appears on the subject's screen until the target was located. Total search time is composed of: system imposed time and user added time. User added time was recorded to evaluate individual search efficiency and to investigate user strategy. The time required to transcribe the information messages is not included in total search time. System imposed time was the estimated minimum amount of time it would require an expert to navigate without error through the database to the target item (Beaudet, 1988).

Expert search time was determined by running a real-time computer simulation of searches for each treatment combination. Each computer simulation run combined the system imposed time requirements plus an additional 0.57 seconds for each menu level selection. This expert time was taken from the American Institutes for Research Data Store (Munger, Smith, and Payne, 1982) for an expert user pressing a push button when cued. System time requirements are composed of three values: system response time (held constant at zero seconds for all conditions), input time-out (manipulated at two and ten seconds), and the amount of time the system requires to speak the menu items. Since the computer operating system and the speech synthesizer vary in the amount of time to respond or to speak the menu items, four simulation runs were used to produce an average expert search time.

Search accuracy. Accuracy of the search strategy was measured by the total number of user key presses. The total number of key presses included essential or expert key presses to locate the target + additional key presses as the result of an inefficient search strategy + any invalid key presses. Expert key presses represent the minimum number of key presses required to complete the target search accurately . Additional key presses are the total number of valid keys used by the subject to arrive at the store item beyond the number of key presses imposed by the system (expert key presses). Invalid key presses occured when subjects pressed keys which were not defined or were inappropriate at that time. Only the "2", "*", "0", and "#" keys were defined for use by the system. Therefore, an invalid key press was recorded whenever any other key was depressed or whenever any one of the defined keys were pressed at the wrong time. For example, pressing the "2" key while searching for the target item was inappropriate and would cause an invalid key press to be recorded.

RESULTS

The data from this study were analyzed in several ways. All five independent variables: speech rate (SR), input time-out (IT), feedback (FB), wallet guide (WG), and database (DB), and the two-way interactions were included in the following analyses. First, an analysis of variance (ANOVA) was conducted on the following search task performance measures: total search time (TST), user added time (UAT), and invalid key presses (IKP). A Newman-Keuls post hoc analysis was performed on the two-way interactions that had a significant effect on total search time and user added time. Regression analysis was conducted to build a linear model for each of these performance measures. A Mann-Whitney U test was conducted to analyze the results from the transcription task. The Mann-Whitney U test was chosen because it is one of the most powerful of the nonparametric tests.

Following the analyses of the performance measures, data collected from the preference measures were analyzed using a nonparametric Mann-Whitney U test. The Mann-Whitney U test was used to determine if any of the independent variables had a significant effect on user preferences. The results of the performance measures and user preferences are reported below in detail.

Performance Measures

Search time. Total search time is composed of system imposed time and user added time. User added time was obtained by subtracting the system imposed time from total search time. An analysis of variance (ANOVA) was conducted on total search time and user added time.

The results of total search time are shown in Table 6; Figure 3 and Figure 4 show the total search time treatment means of main effects and two-way interactions, respectively. The average total search time was 28.53 seconds per trial, with the minimum and maximum total search time of 12.02 and 83.33 seconds, respectively.

The results of the analysis of variance (ANOVA) for total search time show all of the main effects except feedback to have a significant effect on user performance. Speech rate was found to have a significant effect on total search time; those subjects who had 240 words per minute completed each search in an average of 43.29 seconds, while 120 wpm yielded an average total search time of 51.61 seconds per trial. Database Structure also had a significant effect on total search time. As expected, a two second input time-out produced a significantly shorter total search time (25.58 sec) than did a 10 second input time-out (69.32 sec). The availability of a wallet guide also affected search time; when the wallet guide was not available, subjects took longer (51.94 sec) than when they had a wallet guide available (42.96 sec).

Also shown in Table 6, are the three interactions that had a significant effect on total search time. DBxIT was significant, with the shortest total search time (23.59 sec) found when there was a 2x6 database with an input time-out of two seconds, while the longest total search time (84.12 sec) came from the 8x2 database and 10 second input time-out. DBxWG was significant; the shortest total search time (31.21 sec) was acquired with the 2x6 database and the wallet guide, while the longest total search time (56.98) was found when the 8x2 database had no wallet guide. Finally, ITxWG had a significant effect on search time. A two second input time-out with the availability of the wallet guide (27.56 sec) had a shorter total search time than a 10 second input time-out and no wallet guide (76.32 sec).

To determine which conditions of the two-way interactions produced the significant effect on total search time, a Newman-Keuls post hoc analyses was performed on those two-way interactions that had a probability ≤ 0.05 (database structure and input time-out, database structure and wallet guide, and the interaction of input time-out and wallet guide). The results of the post hoc analyses for the two-way interaction of database structure and input time-out showed that there was no difference between a database structure of 2x6 or 8x2 if input time-out is two seconds. On the average, the search task was completed the fastest with an input time out of two seconds, regardless of which database structure was used (see Figure 3). However, with an input time-out of 10 seconds, the type of database structure did make a significant difference on total search time. With the 2x6 database structure and a ten second input time-out, the average total search time was 54.51 seconds, as compared to the 8x2 database structure and a 10 second input time-out which had the worst average total search time of 84.12 seconds.

The results of the post hoc analyses for the two-way interaction of database structure and wallet guide showed that the slowest total search time occurred for the 8x2 database structure, weather a wallet guide was available or not. However, with a database structure of 2x6, the presence of the wallet guide did make a significant difference on total search time. With the 2x6 database structure and no wallet guide available, the total search time on the average was 46.89 seconds, as compared to the 2x6 database structure and a wallet guide available, which had the fastest average total search time of 31.21 seconds.

The results of the post hoc analyses for the two-way interaction of input time-out and wallet guide showed that on the average, the fastest total search time occurred with the 2x6 database structure, weather a wallet guide was available or not. However, with an input time-out of ten seconds, the presence of the wallet guide did make a significant difference on total search time. With input time-out of ten seconds and a wallet guide available, the total search time on the average was 62.32 seconds, as compared to an input time-out of ten seconds and no wallet guide available, which yielded the slowest total search time of 76.32 seconds.

The resulting R -Square value for total search time (with all main effects and all two-way interactions) was 97.18. The R-Square value for the refined, second order, multiple linear regression model was 95.71. This means that approximately 96 percent of the variation in total search time can be explained by the following empirical model:

$$\text{Total Search Time (TST)} = 759.19 - (66.62)SR - (19.62)DB + (349.88)IT - 11.55(WG) + 25.63(DB \times IT) + 53.67(DB \times WG) - (10.04)\, IT \times WG.$$

Table 6
ANOVA for Total Search Time

Source		df	MS	F	p
Treatment (Alias)		(15)			
SR	(DBxITxWGxFB)	1	832.22	29.17	0.0001*
DB	(SRxITxWGxFB)	1	3384.93	118.65	0.0001*
IT	(SRxDBxWGxFB)	1	22953.53	804.57	0.0001*
WG	(SRxDBxITxFB)	1	966.72	33.89	0.0001*
FB	(SRxDBxITxWG)	1	29.12	1.02	0.3200
SR x DB	(ITxWGxFB)	1	32.32	1.13	0.2951
SR x IT	(DBxWGxFB)	1	75.31	2.64	0.114
SR x WG	(DBxITxFB)	1	101.62	3.56	0.0682
SR x FB	(DBxITxWG)	1	81.04	2.84	0.1016
DB x IT	(SRxWGxFB)	1	1970.62	69.07	0.0001*
DB x WG	(SRxITxFB)	1	540.06	18.93	0.0001*
DB x FB	(SRxITxWG)	1	90.61	3.18	0.0842
IT x WG	(SRxDBxFB)	1	302.56	0.61	0.0027*
IT x FB	(SRxDBxWG)	1	18.38	0.64	0.4281
WG x FB	(SRxDBxIT)	1	46.70	1.64	0.2099
Subjects / Treatments		32	28.53		
TOTAL		47			

* Test is significant at $p<0.05$

Factor	Level	Response
Speech Rate (wpm)	240	43.29
	120	51.61
Data Base	8x2	55.85
	2x6	39.05
Input Time-out (sec)	10	69.32
	2	25.58
Wallet Guide	Available	42.96
	Not Avail.	51.94

Figure 3. Summary of Total Search Time: Treatment Means of Main Effects (seconds per trial per subject)

Data Base	Input Time-out (sec): 10	Input Time-out (sec): 2
8x2	84.12	27.57
2x6	54.51	23.59

Data Base	Wallet Guide: Available	Wallet Guide: Not Available
8x2	54.71	56.98
2x6	31.20	46.89

Input time-out (sec)	Wallet Guide: Available	Wallet Guide: Not Available
10	62.31	76.32
2	23.60	27.56

Figure 4. Summary of Total Search Time: Treatment Means for Interactions (seconds per trial per subject)

The results of the ANOVA for user added time are shown in Table 7. Figures 5 and 6 show the treatment means for user added time. The average amount of time added by the user was 7.76 seconds per trial. The minimum amount of user added time was 0.28 seconds per trial, while the maximum user added time was 47.52 seconds per trial.

The ANOVA for user added time found only three main effects (DB, IT, WG) to cause a significant difference. Database differed significantly, with the 8x2 database having the smaller amount of user added time (3.30 sec), while the 2x6 database had an average user added time of 12.21 seconds. This supports previous research done by Beaudet (1988) and Wu (1989). Input time-out caused a difference with a 10 second IT having more user added time (9.62 sec) than a 2 second IT (5.90 sec). Wallet guide also made a difference, with the effects of no wallet guide available significantly increasing user added time (12.26 sec) over its availability, which decreased user added time to an average of 3.27 seconds per trial.

The three interactions which significantly effect user added time are: DBxIT, DBxWG, and ITxWG. The DBxIT interaction showed that either a 2 or a ten second input time-out (3.57 sec, 3.04 sec) with the 8x2 database produced less user added time than a 10 second input time-out with the 2x6 database. The DBxWG interaction found the 2x6 database and no wallet guide available to increase user added time by 20.05 seconds per trial, while the 2x6 database with a wallet guide (4.38 sec), and the 8x2 database with (2.16 sec) or without (4.45 sec) a wallet guide all had similar user added times. The ITxWG interaction showed an input time out of 10 seconds with no wallet guide increased user added time (16.62 sec) over a 10 second input time-out and wallet guide available (2.62 sec) or a 2 second input time out with (3.92 sec) or without (7.89 sec) a wallet guide.

To determine which conditions of the two-way interactions produced the significant effect on user added time, a Newman-Keuls post hoc analyses was performed on those two-way interactions that had a probability ≤ 0.05 (database structure and input time-out, database structure and wallet guide, and the interaction of input time-out and wallet guide). The results of the post hoc analyses for the two-way interaction of database structure and input time-out showed that users added the most time with a ten second input time-out and a 2x6 database structure (15.67 seconds). The other three conditions yielded significantly less user added time (see Figure 6).

The results of the post hoc analyses for the two-way interaction of database structure and wallet guide showed that most of the user added time occurred with no wallet guide available and a 2x6 database structure (20.06 seconds). The other three conditions yielded a significantly less user added time.

The results of the post hoc analyses for the two-way interaction of input time-out and wallet guide showed that most of the user added time occurred with no wallet guide available and a ten second input time-out (16.62 seconds). The other three conditions yielded a significantly faster total search time.

The R-Square value for all main effects and all two-way interactions for user added time was also very high: 79.83 percent. The refined, second order, linear model for user added time had an R-Square value of 67.38, indicating that almost 70 percent of the variation in user added time can be explained by the empirical model:

User Added Time (UAT) = 124.21 - (33.07)DB + (29.74)IT - (11.72)WG - (6.37)DBxIT + (53.56)DBxWG - (10.03)ITxWG.

Table 7

ANOVA for User Added Time

Source		df	MS	F	p
Treatment (Alias)		(15)			
SR	(DBxITxWGxFB)	1	94.93	3.33	0.0775
DB	(SRxITxWGxFB)	1	952.75	33.40	0.0001*
IT	(SRxDBxWGxFB)	1	165.90	5.82	0.0218*
WG	(SRxDBxITxFB)	1	969.01	33.97	0.0001*
FB	(SRxDBxITxWG)	1	27.97	0.98	0.3296
SR x DB	(ITxWGxFB)	1	31.17	1.09	0.3037
SR x IT	(DBxWGxFB)	1	74.05	2.60	0.1170
SR x WG	(DBxITxFB)	1	103.76	3.64	0.0655
SR x FB	(DBxITxWG)	1	79.18	2.80	0.1042
DB x IT	(SRxWGxFB)	1	121.68	4.27	0.0471*
DB x WG	(SRxITxFB)	1	537.82	18.85	0.0001*
DB x FB	(SRxITxWG)	1	86.57	3.03	0.0911
IT x WG	(SRxDBxFB)	1	301.65	10.57	0.0027*
IT x FB	(SRxDBxWG)	1	17.88	0.63	0.4344
WG x FB	(SRxDBxIT)	1	47.30	1.66	0.2071
Subjects / Treatments		32	28.53		
TOTAL		47			

* Test is significant at p<0.05

Factor	Level	Response
Data Base	8x2	3.31
	2x6	12.22
Input Time-out (sec)	10	9.62
	2	5.90
Wallet Guide	Available	3.27
	Not Avail.	12.26

Figure 5. **Summary of User Added Time: Treatment Means of Main Effects (seconds per trial per subject)**

Data Base	Input Time-out (sec): 10	Input Time-out (sec): 2
8x2	3.57	3.04
2x6	15.67	8.77

Data Base	Wallet Guide: Available	Wallet Guide: Not Available
8x2	2.16	4.45
2x6	4.38	20.06

Input Time-out (sec)	Wallet Guide: Available	Wallet Guide: Not Available
10	2.62	16.62
2	3.92	7.89

Figure 6. Summary of User Added Time: Treatment Means for Interactions (seconds per trial per subject)

Search accuracy. The accuracy of the search is described by two dependent measures: invalid key presses, additional key presses. The results of these two performance measures are reported below in detail.

The number of invalid key presses was found to be significantly affected by the availability of the wallet guide. The ANOVA summary for invalid key presses is shown in Table 8. Figure 7 shows the treatment means for invalid key presses. Those with a wallet guide had an average of 0.02 invalid key presses per trial (8 total invalid keys were pressed) as compared to an average of 0.07 invalid keys pressed by those without a wallet guide (26 total invalid key presses). The only interaction found to have a significant difference on the number of invalid key presses is SR x WG. No invalid keys were pressed by the12 subjects in the group which had 120 wpm and a wallet guide available, as compared to 18 invalid keys pressed by the group which had no wallet guide and 240 wpm. The resulting R-Square value for all main effects and all two-way interactions for invalid keys was 38.96 percent. The R-Square value for the refined, second order, linear model was 16.5. This indicates that there is alot of unexplained variation for invalid key presses. The empirical model for invalid key presses is:

$$IKP = 0.71 - (1.50)WG + (0.01)SRxWG.$$

Table 8

ANOVA for Invalid Key Presses

Source		df	MS	F	p
Treatment (Alias)		(15)			
SR	(DBxITxWGxFB)	1	0.0003	0.05	0.8232
DB	(SRxITxWGxFB)	1	0.0081	1.35	0.2562
IT	(SRxDBxWGxFB)	1	0.0003	0.05	0.8175
WG	(SRxDBxITxFB)	1	0.0264	0.05	0.0457*
FB	(SRxDBxITxWG)	1	0.0003	0.48	0.4925
SR x DB	(ITxWGxFB)	1	0.0029	0.48	0.4925
SR x IT	(DBxWGxFB)	1	0.0160	2.62	0.1155
SR x WG	(DBxITxFB)	1	0.0263	4.31	0.0460*
SR x FB	(DBxITxWG)	1	0.0082	1.34	0.2562
DB x IT	(SRxWGxFB)	1	0.0003	0.05	0.8203
DB x WG	(SRxITxFB)	1	0.0003	0.05	0.8203
DB x FB	(SRxITxWG)	1	0.0003	0.05	0.8175
IT x WG	(SRxDBxFB)	1	0.0156	2.62	0.1155
IT x FB	(SRxDBxWG)	1	0.0029	0.48	0.4947
WG x FB	(SRxDBxIT)	1	0.0160	2.61	0.1163
Subjects / Treatments		32	0.0061		
TOTAL		47			

* Test is significant at $p<0.05$

Wallet Guide — Level	Response
Available	0.33
Not Avail.	1.08

Speech Rate (wpm)	Wallet Guide: Available	Wallet Guide: Not Available
240	0.67	0.67
120	0	1.5

Figure 7. Summary of Invalid Key Presses: Treatment Means (key presses per subject)

Pressing valid keys at the incorrect time (i.e., choosing the menu option "fashion" when searching for golf books) would cause the user to press more keys than the required minimum. This measure of search accuracy, Additional Key Presses, yielded two significant main effects: database structure and wallet guide. Two significant interactions were also found: DB x WG, and IT x WG. The results of the ANOVA summary table for additional key presses are shown in Table 9. Figure 8 shows the treatment means for additional key presses.

The 2x6 database increased the number of additional key presses to an average of 0.74 (11.84 for the entire search), as compared to the 8x2 database which had an average additional key presses of only 0.09 (1.46 for the entire search). The lack of a wallet guide also increased the number of additional key presses (0.69 per trial, 10.96 total), as compared to the 8x2 database which had only 0.15 additional key presses per trial (2.34 total).

The interaction of DBxWG found the least number of additional key presses to occur when the 8x2 database was used with a wallet guide (an average of 1.84 additional key presses per subject, or 0.07 additional key presses per trial). There was no difference between database structures, as long as a wallet guide was available. The largest increase in additional key presses occurred when no wallet guide was available and the 2x6 database was used (1.26 key presses per trial). The interaction of IT x WG shows the fewest additional key presses to occur when a 10 second input time out was used with a wallet guide (0.05 key presses per trial), whereas an input time-out of 10 seconds and no wallet guide produced an average of 0.92 additional key presses per trial.

To determine which conditions of the two-way interactions produced the significant effect on additional key presses, a Newman-Keuls post hoc analyses was performed on those two-way interactions that had a probability ≤ 0.05 (database structure and wallet guide, and the interaction of input time-out and wallet guide). The results of the post hoc analyses for the two-way interaction of database structure and wallet guide showed that the most significant amount of user added key presses (20.09) occurred with the 2x6 database structure and no wallet guide available. The other three conditions yielded significantly less user added key presses (see figure 7).

The results of the post hoc analyses for the two-way interaction of input time-out and wallet guide showed that the most significant amount of user added key presses (14.67) occurred with ten second input time-out and no wallet guide available. The other three conditions yielded significantly less user added key presses.

The R-Square value for a linear model of all main effects and all two-way interactions for additional key presses was 62.02. The two-factor interactions linear model built with multiple regression techniques resulted in an R-Square value of 50.4, indicating 50 percent of the variation was explained by the empirical model:

$$AKP = 7.36 - (4.98)DB - (0.98)IT - (0.41)WG + (3.98)DBxWG - (0.71)ITxWG.$$

Table 9

ANOVA for Additional Key Presses

Source		df	MS	F	p
Treatment (Alias)		(15)			
SR	(DBxITxWGxFB)	1	0.0779	0.24	0.6298
DB	(SRxITxWGxFB)	1	4.6488	17.47	0.0002*
IT	(SRxDBxWGxFB)	1	0.2283	0.85	0.3630
WG	(SRxDBxITxFB)	1	4.1219	12.08	0.0001*
FB	(SRxDBxITxWG)	1	0.0360	0.15	0.7019
SR x DB	(ITxWGxFB)	1	0.0686	0.15	0.7019
SR x IT	(DBxWGxFB)	1	0.5612	1.34	0.2552
SR x WG	(DBxITxFB)	1	0.4575	0.92	0.3459
SR x FB	(DBxITxWG)	1	0.5603	2.44	0.1284
DB x IT	(SRxWGxFB)	1	0.5072	1.67	0.2057
DB x WG	(SRxITxFB)	1	2.9671	10.06	0.0033*
DB x FB	(SRxITxWG)	1	0.0098	0.05	0.8287
IT x WG	(SRxDBxFB)	1	1.5287	4.27	0.0471*
IT x FB	(SRxDBxWG)	1	0.0066	0.06	0.8031
WG x FB	(SRxDBxIT)	1	0.0686	0.52	0.4756
Subjects / Treatments		32	0.3273		
TOTAL		47			

* Test is significant at $p<0.05$

Level		Response
Wallet Guide	Available	2.34
	Not Avail.	10.96

Level		Response
Data Base	8x2	1.46
	2x6	11.84

		Wallet Guide	
		Available	Not Available
Data Base	8x2	1.08	1.84
	2x6	3.59	20.09

		Wallet Guide	
		Available	Not Available
Input Time-out (sec)	10	0.92	14.67
	2	3.75	7.25

Figure 8. Summary of Additional Key Presses: Treatment Means (key presses per subject)

Transcription task. The performance measure used to analyze the transcription task was the subject's transcription accuracy score. The ANOVA summary for transcription accuracy, shown in Table 10, indicates that speech rate was the only significant factor. The overall mean transcription accuracy score was 3.43 out of a possible 4.0 words per trial. The average transcription score for 240 wpm was 3.24 (i.e on the average, 82% of the words were transcribed correctly) as compared to 3.62 words for 120 wpm (an average of 90% of the words were transcribed correctly).

User Preferences

Several types of analyses were conducted on user preferences (i.e., F - test, Mann-Whitney U test, t - test) in which users responded to questions about the system using a rating scale from one to seven. Since this study investigated variables set at two levels, the results of the Mann-Whitney U test are reported. The Man-Whitney U test was performed because it is one of the most powerful nonparametric tests. The results of these tests are presented below in detail.

Speech rate was found to have a significant effect on the certainty of transcribing the information message. The results of the preference measures for speech rate are shown in Table 11. As shown in Figure 9, those subjects who received 120 wpm were more certain of their transcription message than those who received 240 wpm (the median value of 240 wpm was 5.75, as compared to a rating of 6.42 for 120 wpm). Logically, speech rate also affected the response to, "how fast did the computer voice speak (one - very slow, seven - very fast) ?" Those who had 120 wpm had an average rating of 3.04, while those with 240 wpm had an average rating of 5.29 (see Figure 10).

The results of the preference measures for input time-out are shown in Table 12. As shown in Figure 11, input time-out was found to affect how users responded to "On a scale of one to seven, how essential was the feedback feature (one - not essential, seven - absolutely essential)?" The mean response for those who had 10 seconds input time-out was 3.67, while those with 2 seconds had an average rating of 5.33. This indicates that the shorter the input time-out, the more essential was feedback. As shown in Figure 12, input time-out also effected how users rated, "How much time did you have to enter a command?" Those subjects with a ten second input time-out had an average rating of 6.58, while those with 2 seconds had an average rating of 4.88 (one- very little, seven - very much).

The database structure of the information system affected several ratings. The results of user preferences for database are shown in Table 13. Figure 13 shows those subjects who used the 8x2 database thought it was less difficult to locate a store item (average rating was 6.96) than did those subjects who used the 2x6 database (average rating was 6.63). Additionally, those subjects with the 8x2 database did not find the wallet guide to be as essential (average rating was 2.83), as did those subjects who used the 2x6 database (average rating was 5.33); see Figure14.

Table 10

ANOVA for Transcription Score

Source		df	MS	F	p
Treatment (Alias)		(15)			
SR	(DBxITxWGxFB)	1	1.6875	25.63	0.0001*
DB	(SRxITxWGxFB)	1	0.0267	0.40	0.5313
IT	(SRxDBxWGxFB)	1	0.0003	0.00	0.9444
WG	(SRxDBxITxFB)	1	0.0000	0.00	1.0000
FB	(SRxDBxITxWG)	1	0.0013	0.02	0.8890
SR x DB	(ITxWGxFB)	1	0.0081	0.12	0.7275
SR x IT	(DBxWGxFB)	1	0.1436	2.18	0.1495
SR x WG	(DBxITxFB)	1	0.0052	0.08	0.7803
SR x FB	(DBxITxWG)	1	0.0208	0.32	0.5777
DB x IT	(SRxWGxFB)	1	0.0117	0.18	0.6759
DB x WG	(SRxITxFB)	1	0.0160	0.24	0.6259
DB x FB	(SRxITxWG)	1	0.2035	3.09	0.0883
IT x WG	(SRxDBxFB)	1	0.0160	0.24	0.6259
IT x FB	(SRxDBxWG)	1	0.0941	1.43	0.2407
WG x FB	(SRxDBxIT)	1	0.0638	0.97	0.3323
Subjects / Treatments		32	0.0658		
TOTAL		47			

* Test is significant at $p<0.05$

Table 11
Summary of User Preferences for Speech Rate

Rating	Z	Prob > IZI
Certainty of transcribing message	3.185	0.001 *
Difficulty of understanding message	1.623	0.105
Difficulty of Locating Store Item	0.274	0.784
Ease of Use of System	-0.198	0.843
Intelligibility of Computer Voice	1.303	0.192
Naturalness of Computer Voice	0.723	0.470
Speech Rate	-5.299	0.000 *
Input time-out	0.956	0.340
Database Structure	-2.19	0.826
Feedback	0.701	0.483
Availability of Wallet Guide	-0.443	0.658

* Test is significant at p<0.05

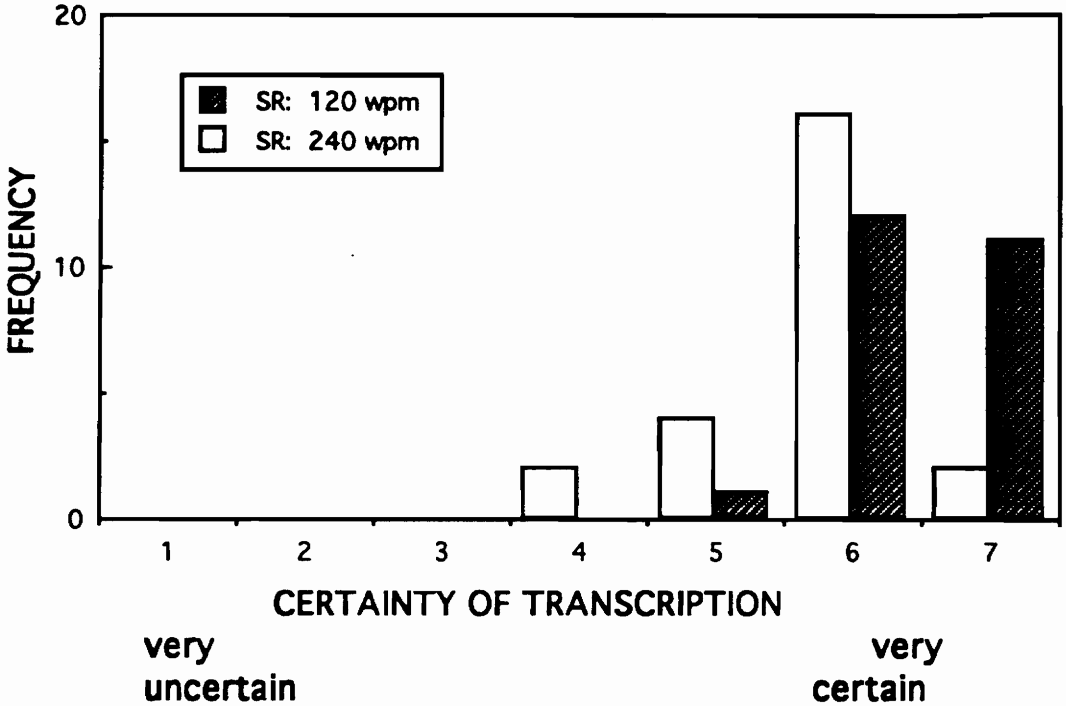


Figure 9. Certainty of Transcription.

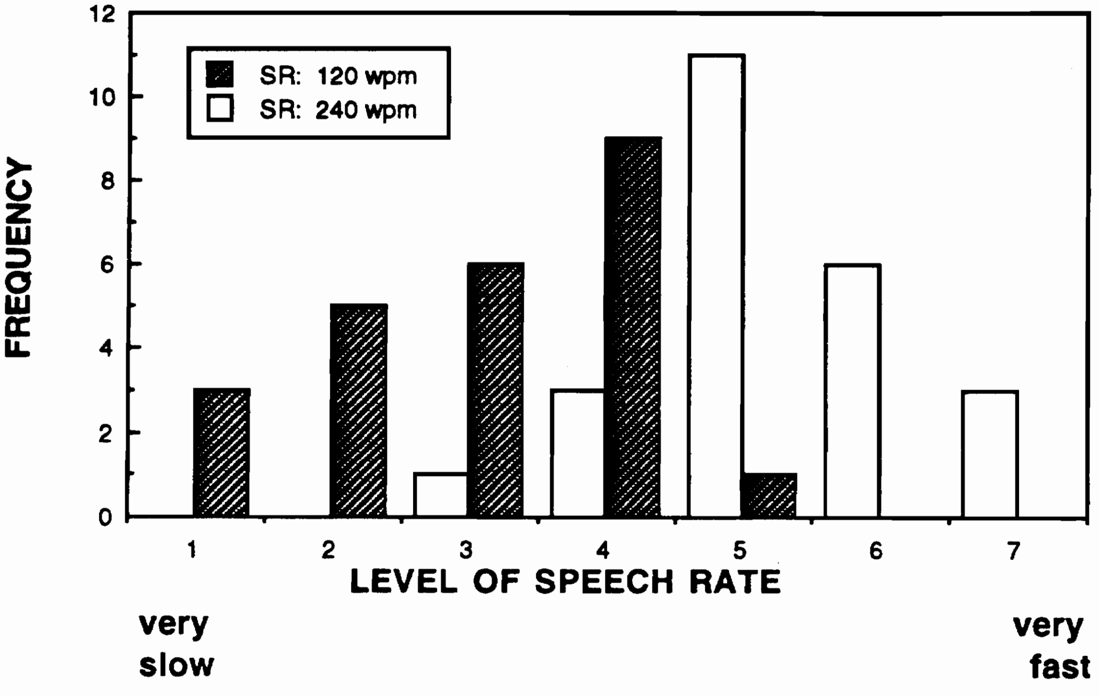


Figure 10. Level of Speech Rate.

Table 12

Summary of User Preferences for Input Time-out

Rating	Z	Prob > IZI
Certainty of transcribing message	-0.408	0.683
Difficulty of understanding message	-0.392	0.695
Difficulty of Locating Store Item	-0.274	0.784
Ease of Use of System	0.341	0.733
Intelligibility of Computer Voice	-1.026	0.305
Naturalness of Computer Voice	0.138	0.890
Speech Rate	0.147	0.883
Input time-out	4.320	0.000*
Database Structure	0.954	0.340
Feedback	-2.267	0.023*
Availability of Wallet Guide	0.354	0.723

* Test is significant at $p<0.05$

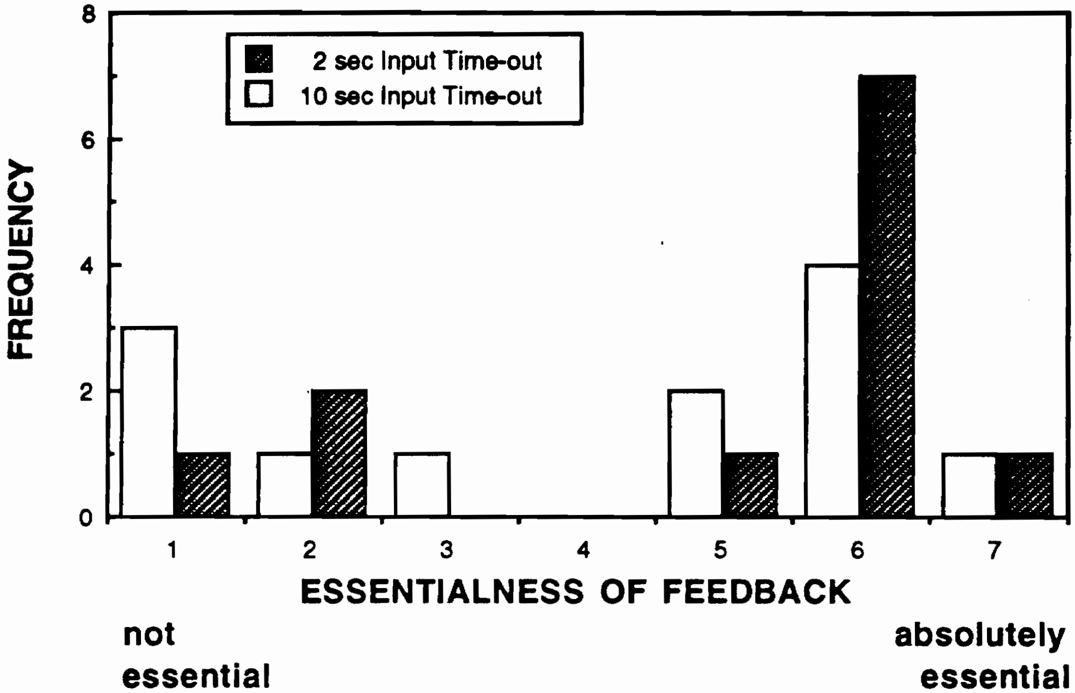


Figure 11. Essentialness of Feedback.

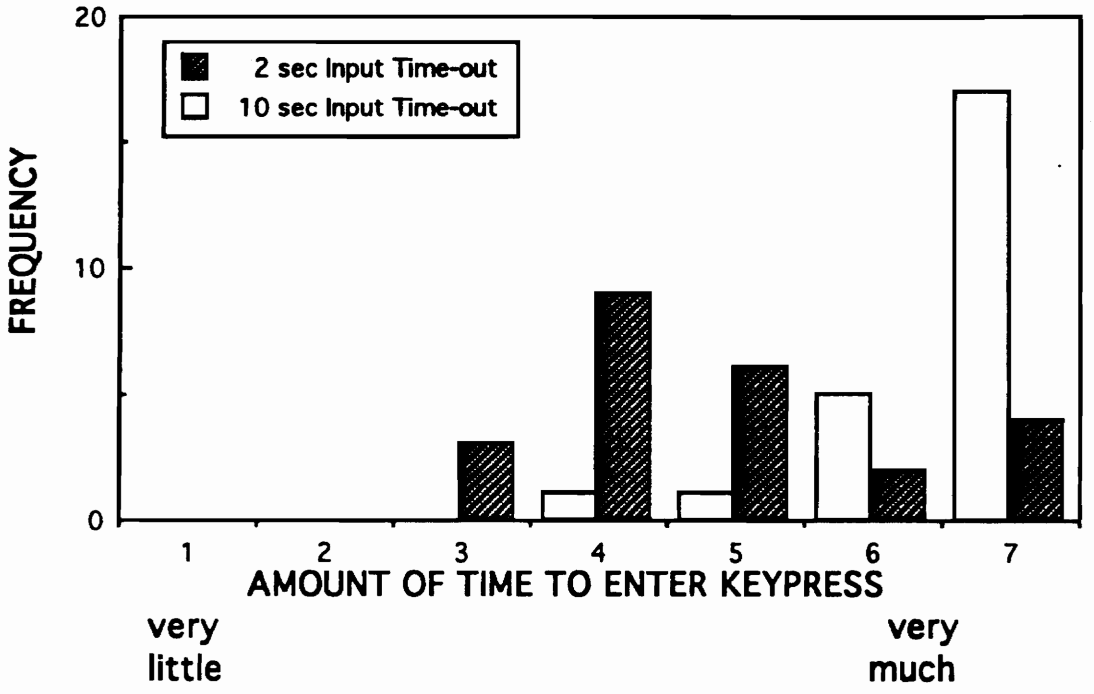


Figure 12. Amount of Time to Enter a Keypress

Table 13

Summary of User Preferences for Database

Rating	Z	Prob > lZl
Certainty of transcribing message	0.128	0.898
Difficulty of understanding message	0.817	0.414
Difficulty of Locating Store Item	-2.560	0.011*
Ease of Use of System	-1.396	0.163
Intelligibility of Computer Voice	-0.598	0.550
Naturalness of Computer Voice	-1.106	0.269
Speech Rate	0.074	0.941
Input time-out	-0.803	0.422
Database Structure	-1.744	0.081
Feedback	-0.031	0.976
Availability of Wallet Guide	2.539	0.011*

* Test is significant at $p<0.05$

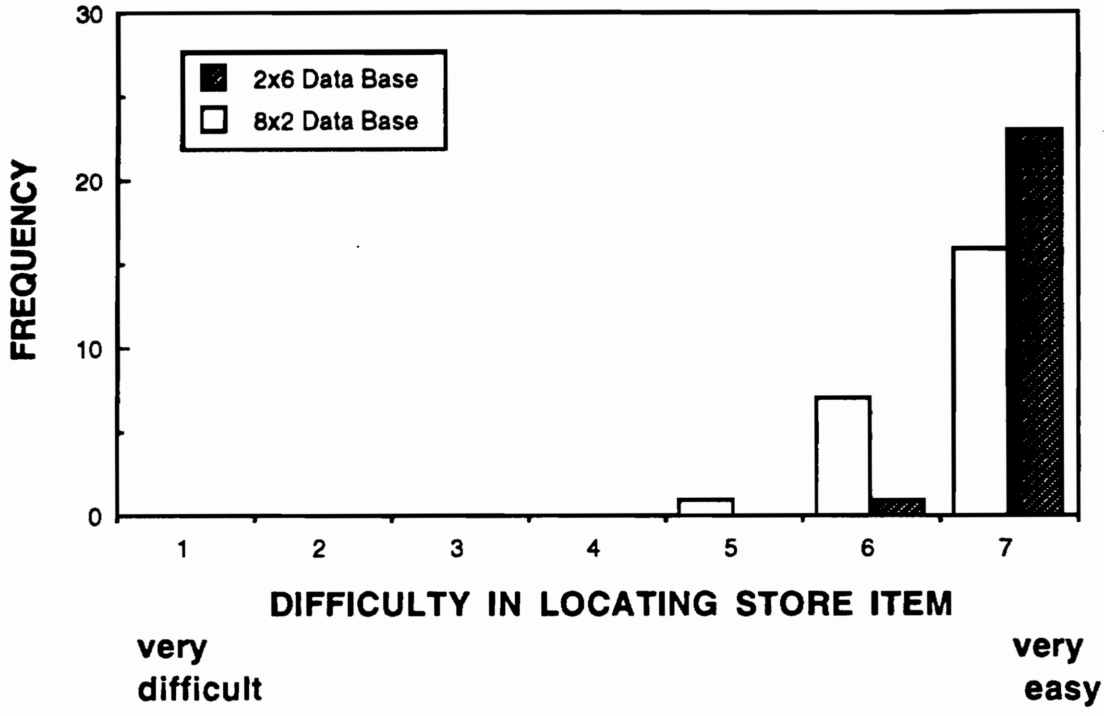


Figure 13. Difficulty in Locating Store Item.

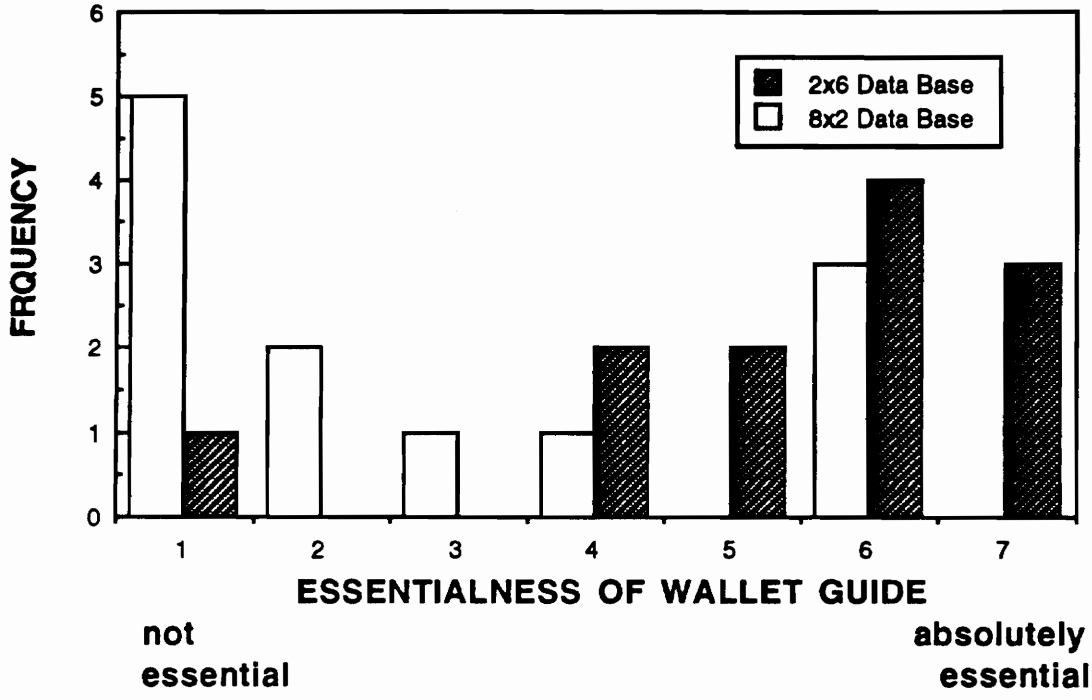


Figure 14. Essentialness of Wallet Guide.

DISCUSSION

The results of this study have important implications for the design of system and user requirements for telephone interfaces. Some of the results are not enlightening, but serve to validate previous research. Other results are unique to this study and, hopefully, will impact future research and the design of telephone interfaces.

Performance Measures

The search task measures were affected by speech rate, input time-out, wallet guide, and database structure, and by the interactions: database structure and input time-out, database structure and wallet guide, input time-out and wallet guide, and speech rate and wallet guide. The results of total search time and user added time were the same except for the effect of speech rate. Total search time was significantly faster when 240 wpm were used. Given the two extreme speech rates used in this study, it is logical that 240 wpm imposed a longer search time than did 120 wpm. This is because total search time includes system imposed time, which is determined by speech rate and input time-out. When speech at 120 wpm was used, it increased the total amount of time required to complete the search.

User added time was less (i.e., subjects completed the search more efficiently) with an 8x2 database structure, a two second linput time-out, or a wallwt guide available. This supports previous research by Kiger (1984), Wu (1989), and Beaudet (1988), respectively.

The effect of the interaction database structure and input time-out showed that an input time-out of two or ten seconds with the 8x2 database structure added much less time to a users search than if a ten second input time-out was used with the 2x6 database structure. The interesting result is that the combination of two seconds input time-out with the 2x6 database structure was not significantly different from the results of the 8x2 database structure at either input time-out. This indicates that long input time-outs and a complex database structure combine to confuse users. It is possible that the long input time-out hindered working memory, thereby forcing users to listen to the menu options several times before making the correct selection. This result indicates that when designing a telephone interface for large information systems, the designer should avoid long input time-out levels (greater than ten seconds).

The database structure and wallet guide interaction suggest an important finding. Although this research and previous studies have substantiated the superiority of the 8x2 database structure, the availability of a wallet guide enhanced user performance with the more complex, 2x6 database structure. The performance of the subjects using the wallet guide with the 2x6 database structure was as fast as those using the 8x2 database structure with or without the wallet guide. Since wallet guides do not hinder performance with a simple database, but enhance user performance when using complex database structures, wallet guides should be provided whenever possible in the design of telephone interfaces.

The input time-out and wallet guide interaction showed an important result with respect to user added time. A ten second input time-out and no wallet guide hinder the ability of the user to recall what menu options had been listed. The most beneficial combination is a long input time-out with a wallet guide. This reduced user added time to a minimal 2.62 seconds. This indicates that if a wallet guide is provided, the system should allow the user enough time to look at, read, and utilize it.

The accuracy of the search depended on two measures: the number of additional keys that were used, and the number of invalid keys pressed. As expected, the availability of the wallet guide significantly reduced the number of invalid and additional key presses. This indicates that telephone information retrieval is more accurate when users have access to a wallet guide.

The 8x2 database structure reduced the number of times additional keys and invalid keys were pressed. However, the results from the interaction of database structur and wallet guide indicate that the availability of the wallet guide can overcome the complexities of searching thru a 2x6 database structure. Even though the fewest additional keys were pressed when the 8x2 database was used with a wallet guide, there was no significant difference between the 8x2 or 2x6 database structure as long as the wallet guide was used. The 2x6 database structure used without a wallet guide did cause a significant increase in the number of additional key presses.

The significance of the linput time-out and wallet guide interaction when measuring additional key presses reinforces the importance of having a wallet guide available. If a wallet guide is not provided, and a ten second input time-out is imposed (the worst combination for short term memory), the number of additional keys pressed is significantly higher than when a ten second input time-out was used with a wallet guide available. This also suggests the importance of providing enough time for the user to read the wallet guide.

Finally, the number of times invalid keys were pressed was significantly reduced when the speech rate was 120 wpm and a wallet guide was available, as compared to 240 wpm used without a wallet guide. The result that speech rate of 120 or 240 wpm did not differ as long as a wallet guide was available, is due to the fact that the wallet guide aided the user in understanding the synthetic speech. Words spoken at high speech rates (which may have been unintelligible) were read by the user when a wallet guide was available, thereby removing any uncertainty.

Although feedback failed to have a significant effect on search time or search accuracy, its inclusion to the design of telephone interfaces is still recommended. Even though the conditions of feedback were "available" and "not available," a certain amount of feedback was always present - based on the design of this telephone interface. The levels of feedback should be thought of as "feedback available" and "additional feedback available." For example, menu options were automatically repeated once the list of options had been exhausted. Also, by pressing the "*" key, users could hear previous menu options repeated. Since some amount of feedback was inherent to the system, the results of this study should be interpreted as, the additional amount of feedback offered did not have a significant effect on user performance.

User Preference

The subjective ratings used to evaluate user preferences did show the importance of feedback varied with the amount of input time-out imposed by the system. When a two second input time-out was imposed, subjects reported that feedback was more essential than when a ten second input time-out was used. This is understandable considering that when subjects were given only two seconds to enter a command, they were more likely to "miss" the menu option that they wanted to choose. Thus, feedback allowed them to hear which menu option they actually selected.

Subjective ratings were also significantly effected by the structure of the information system. Subjects who used the 8x2 database structure reported that they were able to locate store items easier than those subjects who had the 2x6 database structure. Also, subjects who used the 8x2 database structure did not feel the wallet guide was essential to complete the search task. However, those who had the 2x6 database structure ranked the wallet guide as being very essential. These results substantiate previous findings from this study.

Further Research

The extreme levels of the variables used in this study were chosen to aid future data collection process of the sequential research paradigm. At the onset of this study, it was thought that these extreme values would simplify the data-bridging and optimization procedures to be performed. In order to predict optimum responses within the range of the design space, the extreme values of this space must be investigated. Therefore, this study provides the building blocks for the third stage of the sequential research paradigm as described by Williges and Williges (1989).

However helpful the extreme levels of the independent variables will be to the optimization stage, many suggestions for further research are due to the extreme values used in this study. Although there are no other levels available for wallet guide (e.g., it is either available or not available), feedback, input time-out, speech rate, and database structure can be manipulated at many other levels. Further research should be done to investigate the usefulness of a 3x4 or 4x3 database structure. These structures would provide comparable results since they offer the same number of end nodes as the 2x6 and the 8x2 database structures. Also, to simulate the "real world" more closely, unsymmetrical database structures should also be evaluated.

CONCLUSIONS

Telephone Information System

The results of this study indicate that when designing a telephone information system:

1. A graphical representation of the database structure should always be provided, especially when the database is complex (i.e., complicated databases will not degrade user performance if a WG is available).
2. When a WG is provided, the IT level should be long enough to allow the user time to read, understand, and utilize it.
3. SR can be increased without degrading search performance, if the WG is provided.
4. If it is impossible to offer a WG, then long input time-outs should be avoided.
5. The 8x2 DB enhances user performance by allowing for a faster and more accurate search. If a WG cannot be provided, the 8x2 DB is the recommended structure to use.
6. When IT is short, more FB should be offered, especially to inform the user which option has been selected.

Sequential Experimentation

The results of this study indicate that when designing studies for sequential research:

1. A Resolution V design enhances the researcher's understanding of the variables identified in the Resolution IV design used in the

screening study. Specifically, this study not only reinforced the screening study, that five variables significantly affected human performance, but also described the relationship among those five variables, and how their interactions affected human performance.

2. The 2^k fractional factorial experimental designs are useful when the amount of data collection required by the complete factorial design is more than the experimenter can afford.
3. The 2^k fractional factorial designs are useful for fitting regression models when k is large and higher-order interactions are not considered to be important.
4. For k=5, the reduction in the number of data points is accomplished at the expense of three, four, and five-way interactions while the main effects and two-way interactions are still independently estimated.
5. This design is especially applicable to human factors research since, generally, high order interactions do not contribute to the explanation of human performance in complex systems.
6. When using fractional factorial designs, it is important to choose the defining relationship carefully, so that main effects and two-way interactions are not confounded with each other.
7. The more fractionated a design is, the more information is sacrificed in the experiment. Researchers must not reduce the size of the experiment too much, or they may defeat the original purpose of the study.

REFERENCES

Beaudet, D.B. and Williges, R.C. (1988). The role of screening studies in sequential research designs. *Proceedings of the Human Factors Society 32nd annual meeting.* Santa Monica, CA: Human Factors Society, 1174-1178.

Billingsly, P.A. (1982). Navigation through hierarchical menu structures: does it help to have a map? *Proceedings of the Human Factors Society 26th annual meeting.* Seattle, WA: Human Factors Society, 103-107.

Diamond, W. J. (1981). *Practical experiment designs for engineers and scientists.* Belmont, CA: Lifetime Learning.

Elmer-Dewitt, P. (1989). Hello! This is voice mail speaking. *Time, 133* , 98.

Gould, J. D., and Boies, S. J. (1984). Speech filing - An office system for principles. *IBM Systems Journal, 23/1*, 65-81.

Greene, B. G., Manous, L. M., and Pisoni, D. B. (1984). Perceptual evaluation the DECtalk: a final report of version 1.8. In *Research on Speech Perception Report No 10* (pp. 77-127). Bloomington, IN: Indiana University.

Halstead-Nussloch, R. (1989). The design of phone-based interfaces for consumers. (Tech. Report D59A/701-S). Poughkeepsie, NY: IBM Human Factors Development Lab.

Herlong, D. W. and Williges, B.H. (1988). Designing speech displays for telephone information systems. *Proceedings of the Human Factors Society 32nd annual meeting.* Santa Monica, CA: Human Factors Society, 215-218.

Holander, M. and Wolfe, D. A. (1973). *Non-parametric statistical methods.* New York: John Wiley & Sons.

Keppel, G. (1982). Design and Analysis: A Researcher's Handbook. Pentice-Hall, Inc., New Jersey.

Kidd, A. L. (1982). Problems in man-machine dialogue design. In *IEEE Proceedings of Sixth Conference on Computer Communications* (pp. 531-536). North-Holland, Amsterdam: North-Holland.

Kiger, J. I. (1984). The depth/breadth trade-off in the design of menu-driven user interfaces. *International Journal of Man-Machine Studies, 20,* 201-213.

Merva, M. A. (1987). *The effects of speech rate, message repetition, and information placement on synthesized speech intelligibility.* Unpublished masters thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

MIller, D. P. (1981). The depth/breadth trade-off in hierarchical computer menus. *In Proceedings of the Human Factors Society 25th Annual Meeting* (pp 296-300). Santa Monica, CA: Human Factors Society.

Munger, S., Smith, R. W., and Payne, D. (1962). *An index of electronic equipment operability: Data store.* Pittsburgh: American Institutes for Research.

Myers, R. H. (1976). Response Surface Methodology. Virginia Polytechnic Institute and State University.

Podgorny, P. (1985). Telephone as computer terminal. In *The Official Proceedings of Speech Tech `85* (pp. 103-109). New York: Media Dimensions.

Schmandt, C. (1985). Voice access to an electronic mail system. In *The Official Proceedings of Speech Tech `85* (pp. 89-91). New York: Media Publications.

Simon, C.W. (1973). Economical multifactor designs (Tech. report P73-326). Culver City, CA: Hughes Aircraft Co.

Simon, C. W., and Roscoe, S. N. (1984). Application of a multifactor approach to transfer of training research. *Human Factors, 26*, 591-612.

Simpson, C. A., and Marchionda-Frost, K. (1984). Synthesized speech rate and pitch effects on intelligibility of warning messages for pilots. *Human Factors, 26*, 509-517.

Snowberry, K., Parkinson, S. R., Sisson, N. (1983). Computer display menus. *Ergonomics, 26 (7),* 699-712.

Tatro, J. S., and Roscoe, S. N. (1986). An integrated display for vertical and translational flight: Eight factors affecting pilot performance. *Human Factors, 28*, 101-120.

Waterworth, J., and Lo. A. (1984). Examples of an experiment: Evaluating some speech synthesizers for public announcements. In A. Monk (Ed.), *Fundamentals of human-computer interaction.* London: Academic Press

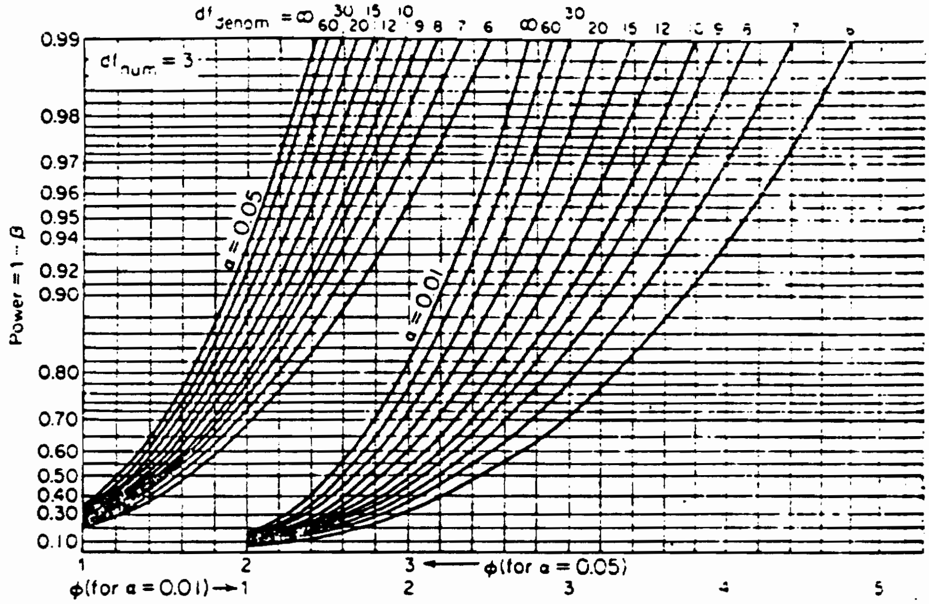
Williges, R.C., and Williges, B.H. (1989). Integrated research paradigm for complex experimentation. Proceedings of the Human Factors Society 32nd annual meeting. Denver, CO: Human Factors Society.

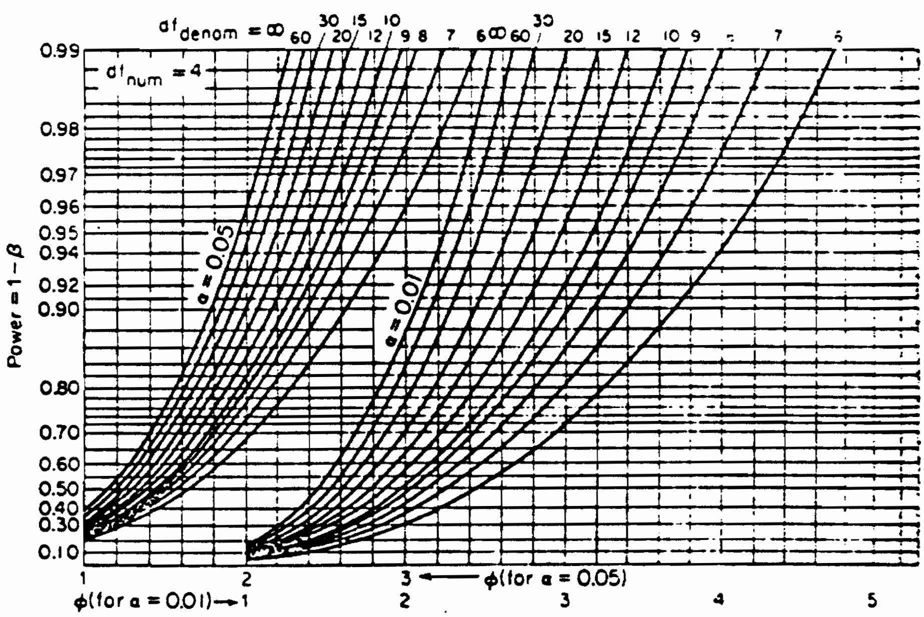
Witten, I. H., and Mandams, P. H. C. (1977). The telephone inquiry service: A man-machine system using synthetic speech. *International Journal of Man-Machine Studies, 9*, 449-464

Wu, J.K.K (1989). *Assessing design trade-offs in a telephone-based information system.* Unpublished masters thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

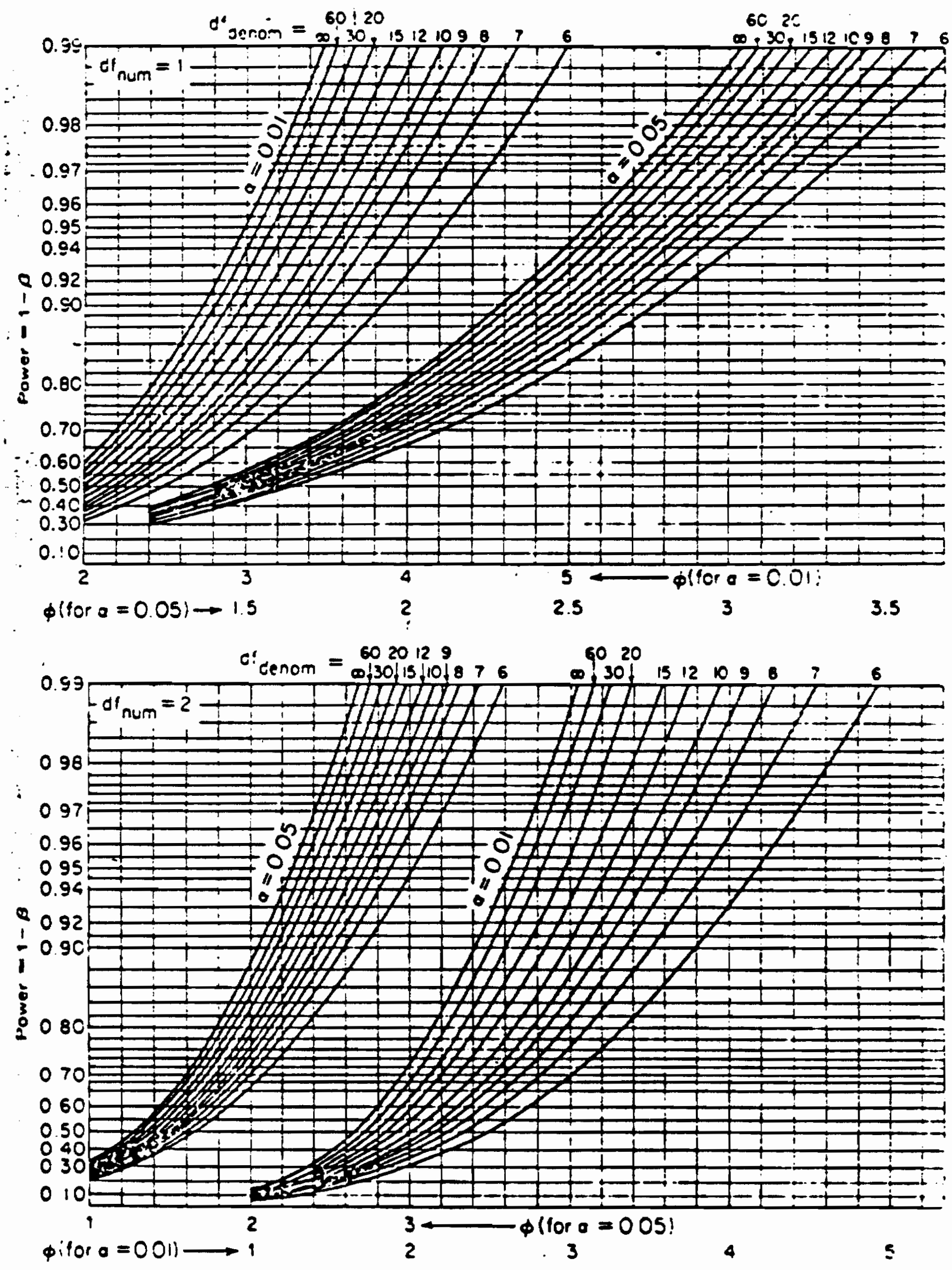
Appendix I

Power Analysis





Power Functions for Analysis of Variance (from Keppel).



Power Functions for Analysis of Variance (from Keppel).

Appendix II

Subject's Instructions

Your task is to search for information on store items in the department store's talking database. Store items will be presented as targets on the computer display in front of you. You will find the target by using the telephone keys to move through the talking database.

These are your instructions:

1. Press the ON/OFF key on the telephone keypad and listen for a dialtone.
2. Press the DIAL key on the telephone keypad (upper right corner).
3. The talking computer will answer the telephone and offer you instructions. Press the **# key** on the telephone keypad and listen carefully to the instructions for using the telephone keypad.
4. Read the first target on the computer display in front of you.
5. Watch the computer display. It will signal you when the search is about to begin.
6. The talking computer will begin speaking a menu of keywords.Keywords categorize groups of store items. After each keyword is spoken, the computer will pause briefly to allow you to select the item.If you do not select the item, the computer will speak another keyword for that menu.
7. To locate the target, select a keyword from the menu which best categorizes the store item you are searching for. The computer will then speak a new menu of keywords, based on your selection. If you need to hear the keypad instructions again, select HELP from any menu.
8. Continue listening to menus and selecting keywords until you reach the desired store item.
9. When you hear the desired store item, press the **2 key** on the telephone keypad and listen carefully to the information message.

10. The computer display will prompt you to transcribe what you heard.
11. Type the information message you heard into the computer, and press the RETURN key.
12. Rate the certainty of your transcription being correct on a scale of 1 (very uncertain) to 7 (very certain), and press the RETURN key.
13. Rate the difficulty of understanding the message on a scale of 1 (very difficult) to 7 (very easy), and press the RETURN key.
14. Rate the difficulty of locating the store item on a scale of 1 (very difficult) to 7 (very easy), and press the RETURN key.
15. Read the next target on the computer display and get ready to start the next search. The computer display will signal you to begin the next search and will speak the first item in the main menu. Locate the next target and transcribe the information message.
16. The experiment will proceed in this fashion. You will search for a total of 16 targets.
17. The computer display will indicate when you have completed the target searches. The computer display will then request that you rate certain characteristics of the telephone information system. The meaning of each characteristic and how it should be rated will be explained on the computer display.

If you have any questions, please ask the experimenter now.

Appendix III

Informed Consent Form

Participant's Informed Consent Form

The following experiment is a study concerning the evaluation of a telephone based information system. During the experiment, you will be monitored with a closed circuit video system. As a participant in this experiment, you have certain rights as explained below. The purpose of this document is to describe these rights and to obtain your written consent to participate in the experiment.

1. You have the right to discontinue your participation in the study at any time for any reason. If you decide to terminate the experiment, inform the researcher and he will pay you for the length of time you have participated.

2. You have the right to inspect your data and withdraw it from the experiment if you feel that you should for any reason. In general, data are processed and analyzed after a subject has completed the experiment. At that time, all identification information will be removed and the data treated with anonymity. Therefore, if you wish to withdraw your data, you must do so immediately after your participation is completed.

3. You have the right to be informed of the overall results of the experiment. If you wish to receive a synopsis of the results, include your address with your signature below. If after receiving the synopsis, you would like more in depth information, please contact Virginia Tech's Human-Computer Interaction Laboratory and a full report will be made available to you.

This research is funded by a research contract with the National Science Foundation. The co-principal investigators are Dr. Robert Williges, and Ms. Beverly Williges. The researcher is Michele M. Martin. All of these people can be contacted at the following address and phone number:

Human-Computer Interaction Laboratory
302 Whittemore Hall
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
(703) 231-4602

Further comments or questions can be addressed to Charles Waring, chairman of the Institutional Review Board for the Use of Human Subjects in Research. He can be contacted at the address and the phone number listed below:

Mr. Ernest R. Stout
Associate Provost for Research-Office of Sponsored Research Programs
301 Burruss Hall
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
(703) 231-5281

If you have any questions about the experiment or your rights as a participant, please do not hesitate to ask. The researcher will do his best to answer them, subject only to the constraint that he does not pre-bias the experimental results.

Your signature below indicates that you have read and understand your rights as a participant (as stated above), and that you consent to participate.

Participant's Signature

Witness's Signature

Print name and address if you wish to receive a summary of the experimental results.

Appendix IV

Questionnaire

Subject number _____

Subject Information Questionnaire

Age: _____ Sex: _____ Native language: _________
(include region of rearing)

Please list any hearing impairments you may have.

____

For the following questions, please circle the most accurate response.

- How experienced are you with using computers?

no experience	some experience	experienced	very experienced

- How experienced are you with using information systems?

no experience	some experience	experienced	very experienced

- How experienced are you with listening to synthesized speech?

no experience	some experience	experienced	very experienced

Appendix V

Subjective Rating Scales

Subjective ratings collected for each store item and information message transcription.

Certainty in Transcribing Information Message

1-----2-----3-----4-----5-----6-----7
very uncertain ... very certain

On a scale of 1 to 7, how CERTAIN are you of your transcription?

Difficulty of Understanding Information Message

1-----2-----3-----4-----5-----6-----7
very difficult ... very easy

On a scale of 1 to 7, how DIFFICULT was it to understand the message?

Difficulty of Locating Store Item

1-----2-----3-----4-----5-----6-----7
very difficult ... very easy

On a scale of 1 to 7, how DIFFICULT was it to locate the store item?

Subjective ratings collected after all the target searches are completed.

Ease-of-use

1-----2-----3-----4-----5-----6-----7
very difficult ... very easy

On a scale of 1 to 7, how easy was the information system to use?

Computer Voice Intelligibility

1-----2-----3-----4-----5-----6-----7
very unintelligible — very intelligible

On a scale of 1 to 7, how intelligible was the computer voice?

Computer Voice Naturalness

1-----2-----3-----4-----5-----6-----7
very unnatural — very natural

On a scale of 1 to 7, how natural was the computer voice?

Computer Voice Speech Rate

1-----2-----3-----4-----5-----6-----7
very slow — very fast

On a scale of 1 to 7, how fast did the computer voice speak?

Input Time-out

1-----2-----3-----4-----5-----6-----7
very little — very much

On a scale of 1 to 7, how much time did you have to enter a command?

Database Organization

1-----2-----3-----4-----5-----6-----7
very complex — very simple

On a scale of 1 to 7, how complex was the organization of the database?

Wallet Guide

1-----2-----3-----4-----5-----6-----7
not essential — absolutely essential

On a scale of 1 to 7, how essential was the wallet guide feature?

Feedback

1-----2-----3-----4-----5-----6-----7
not essential — absolutely essential

On a scale of 1 to 7, how essential was the feedback feature?

Appendix VI

Targets and Information Messages

The following are store items and associated information messages. Information messages are classified as being an availability (A), information (I), location (L), or price (P) oriented messages.

Recliner Chairs: Leather coverings are offered to wholesale buyers. (I)

Hope Chests: Walnut stains are reduced by 34 to 40%. (P)

Washers: Deluxe models are available with green trimming. (A)

Food Blenders: Boxes and cartons are in the wrapping center. (L)

Guitars: Carrying cases are reduced by 55 to 63%. (P)

Compact Discs: Head cleaners are on aisle 12. (L)

Chicken Cookbooks: Collector editions are available in limited quantities. (A)

Football Books: Faculty discounts are offered to gym teachers. (I)

Men's Blazers: Garment bags are offered with new purchases. (I)

Men's Sweaters: Rugby letters are sold for $11.60. (P)

Knit Dresses: Designer collections are available in red and ivory. (A)

Silk Blouses: Maternity wear is near ladies lingerie. (L)

Gold Chains: Instant financing is available at the central office. (A)

Pearl Necklaces: Sorority clasps are in the school department. (L)

Eye Mascara: Travel supplies are sold for $17.50. (P)

Oriental Fragrances: Manufacturer's samplers are offered to interested shoppers.(I)

VITA

Michele Marie Cary

Bell Northern Research/Northern Telecom
Dept 6D21
35 Davis Drive
Research Triangle Park, NC 27709
(919)-991-8769

Education

Master of Science, Industrial Engineering and Operations Research - Human Factors, Virginia Polytechnic Institute and State University, Blacksburg, Virginia. Thesis: "Optimizing the Usability of a Telephone-Based Information System"
QCA: 3.8 / 4.0

Bachelor of Science, Industrial Engineering and Operations Research, Virginia Polytechnic Institute and State University, Blacksburg Virginia. May 1988

Employment Experience

Member of Scientific Staff: BNR June 1990 - Present
Systems Engineer responsible for ISDN and Analogue Services on DMS-100.

Graduate Teaching Assistant: VPI&SU January 1990 - June 1990
Course: IEOR 5616, Human Factors Design Methodology II. Responsible for preparing and presenting course lectures for SAS and STATVIEW-512, as well as teaching, preparing, and grading all tests and homework assignments.

Research Assistant: Human Computer Interaction Laboratory
August 1988 - December 1989
Employed as a team member to conduct research on computer enhanced communications systems.

Systems Analyst: Philip Morris, USA System Design and Development.
May1988 - August 1988
Major project focused on the conceptual design, development, and implementation of a prototype Management Information System (MIS) for use by the Plant Manager.

Computer Programmer: Philip Morris, USA System Design and Development
May 1987 - August 1987
Responsible for supporting management with operational research type reporting through SAS and FOCUS. Major project included developing statistical quality control charts for the cigarette manufacturing center.

Research Experience

Independent Research Project: January 1988 - June 1988
Responsible for designing, conducting, and analyzing the results of an experiment concerning the effects of computer generated speech rate on transcription accuracy.

Senior Design Project: September 1987 - May 1988
Analyzed the current manufacturing expediting system at Poly-Scientific and suggested appropriate changes. Proposal gained approval at all levels of management.

Publications

Martin, M.M., Williges,R.C., and Williges, B. H. (1990). Improving the usability of a telephone-based information system. Submitted for publication in *Proceedings of the 34th Annual Meeting of the Human Factors Society*. Human Factors Society, Orlando, FL.

Activities/Awards

Scholastic Honors

- Alpha Pi Mu: Industrial Engineering Honor Society
- Gamma Beta Phi: Honor and Service Organization
- Outstanding College Students of America
- Secretary, Human Factors Society, Student Chapter - VPI&SU
- Alumni Committee Chairman, Human Factors Society, Student Chapter
- Speaker Presentation Committee, Human Factors Society, Student Chapter-VPI&SU

Athletic Honors

- Full Athletic Scholarship
- Varsity Swimteam
- Intramural Volleyball Co-captain

Professional Organizations

- Human Factors Society
- Association for Computing Machinery
- Computer-Human Interaction Special Interest Group
- Artificial Intelligence Special Interest Group
- Society of Women Engineers

Michele Marie Cary