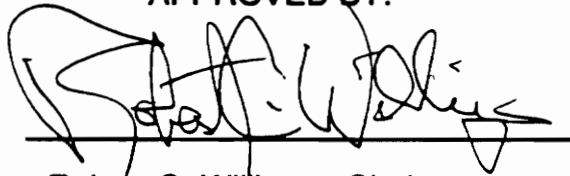
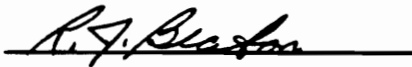


**Integrated Empirical Models
Based on a Sequential Research Strategy**

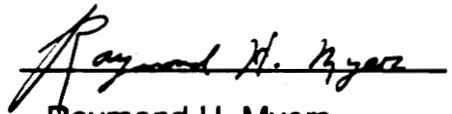
by
Sung Ho Han

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Industrial and Systems Engineering

APPROVED BY:


Robert C. Williges, Chairman

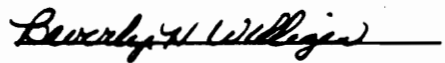
Robert J. Beaton



Raymond H. Myers



Dennis L. Price



Beverly H. Williges

January, 1991
Blacksburg, Virginia

c.2

LD
5655
V856
1991
H37
c.2

Integrated Empirical Models
Based on a Sequential Research Strategy

by
Sung Ho Han

Committee Chairman: Robert C. Williges
Industrial and Systems Engineering
(ABSTRACT)

A systematic research approach is necessary to investigate complex systems. This approach should provide a tool for examining multifactors in an efficient manner since a large number of factors is usually involved in the design and evaluation of complex systems. This study is used to develop empirical models which describe the functional relationships of many independent variables in the design of a telephone information system. Such a development is based on integrating several data sets using sequential experimentation.

Reanalyses of previous experiments were conducted to examine necessary and sufficient conditions for integrating data sets resulting from previous studies. As a result, an experiment was conducted to investigate the effects of independent variables which were not manipulated in the previous experiments. An additional experiment was conducted to provide a bridge among several data sets. The integrated data set was then used to build second-order empirical models using polynomial regression. Determinant values of $X'X$ matrices served as a statistical criterion for achieving minimum

variances of coefficients and prediction variances of the models. Based upon the empirical models developed, optimum configurations of the telephone information system were obtained using a nonlinear programming technique. A separate optimization method was used since the empirical models included both continuous variables and discrete variables.

Specific procedures and guidelines are suggested in planning and conducting sequential research which deals with a large number of independent variables in an efficient and systematic manner. The procedures and guidelines are summarized based upon the lessons learned from the dissertation research. These include administrative requirements, alternative experimental designs, methodological considerations on conducting sequential experiments, and other necessary rules and decision criteria for bridging data sets and optimizing empirical models. This approach is expected to provide a tool for obtaining generalizable results in human factors research.

**Integrated Empirical Models
Based on a Sequential Research Strategy**

by

Sung Ho Han

Committee Chairman: Robert C. Williges

Industrial and Systems Engineering

(ABSTRACT)

A systematic research approach is necessary to investigate complex systems. This approach should provide a tool for examining multifactors in an efficient manner since a large number of factors is usually involved in the design and evaluation of complex systems. This study is used to develop empirical models which describe the functional relationships of many independent variables in the design of a telephone information system. Such a development is based on integrating several data sets using sequential experimentation.

Reanalyses of previous experiments were conducted to examine necessary and sufficient conditions for integrating data sets resulting from previous studies. As a result, an experiment was conducted to investigate the effects of independent variables which were not manipulated in the previous experiments. An additional experiment was conducted to provide a bridge among several data sets. The integrated data set was then used to build second-order empirical models using polynomial regression. Determinant values of $X'X$ matrices served as a statistical criterion for achieving minimum

variances of coefficients and prediction variances of the models. Based upon the empirical models developed, optimum configurations of the telephone information system were obtained using a nonlinear programming technique. A separate optimization method was used since the empirical models included both continuous variables and discrete variables.

Specific procedures and guidelines are suggested in planning and conducting sequential research which deals with a large number of independent variables in an efficient and systematic manner. The procedures and guidelines are summarized based upon the lessons learned from the dissertation research. These include administrative requirements, alternative experimental designs, methodological considerations on conducting sequential experiments, and other necessary rules and decision criteria for bridging data sets and optimizing empirical models. This approach is expected to provide a tool for obtaining generalizable results in human factors research.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation on contract number IRI 8604793. Dr. Laurence C. Rosenberg served as the technical monitor. I would like to thank Dr. Robert C. Williges and Ms. Beverly H. Williges for their guidance throughout this research. Thanks goes to Dr. Dennis L. Price and Dr. Robert J. Beaton for their comments. I am truly indebted to Dr. Raymond H. Myers for his insightful contributions and recommendations which greatly enhanced this study. Thanks also goes to Mr. Ben E. Selbey and Dr. Walter W. Wierwille for their assistance in recruiting subjects; Dr. Hanif D. Sherali for his comments on optimization techniques; Joseph D. Chase for his contribution to the development of a data reduction software.

Especially, I want to thank my mother and grandmother for their unlimited support and my father-in-law for his encouragement. Most of all, I want to thank my lovely wife Youngmae, who always gave me motivation and conviction.

DEDICATION

To my little Chungmin.

TABLE OF CONTENTS

List of Illustrations	xii
List of Tables	xiv
Glossary	xviii
CHAPTER 1: INTRODUCTION	1
Design and Evaluation of Human-Computer Interfaces	1
Design Procedures	1
Research Methods	3
Approaches in usability engineering	5
Laboratory experimentation	6
Problems of human factors research	7
Empirical Models as Design and Evaluation Tools	8
Quantitative Models	9
Sequential Research Strategy as a Tool for Empirical	
Model building	10
Selection	11
Description	11
Optimization	13
Research Objectives	13
Approaches to Achieving the Objectives	14
Primary objective	14
Secondary objective	15

CHAPTER 2: TELEPHONE INFORMATION SYSTEMS	17
Background	17
Hypothetical Telephone Information System	22
Characteristics of the Hypothetical System	23
Potential factors of interest	28
Initial Selection of Variables	28
Initial selection of independent variables	30
Screening of initial variables	31
CHAPTER 3: REANALYSIS OF PREVIOUS STUDIES	34
Background	34
Feasibility Analysis	34
Experimental Space	35
Independent variables	35
Experimental space	35
Analysis Procedures	36
Redefinition of Dependent Measures	36
Analysis Methods	37
Results of Reanalysis	38
Screening Study	38
Experiments for Description of Independent	
Variables	39
Experiment 1	44
Experiment 2	44
Experiment 3	47

CHAPTER 4: EXPERIMENT 3	50
Method	50
Experimental Design	51
Subjects	53
Experimental Tasks	56
Apparatus	56
Speaker phone	56
Speech synthesizer	57
Computer system	57
Audio/video equipment	58
Experimental rooms	58
Information Databases	58
Structures of databases	60
Product information messages	61
Experimental Procedures	61
Instruction session	61
Practice session	65
Experimental session	65
Post experimental session	68
Dependent Measures	69
Results	69
Total Search Time (TST)	69
User Added Time (UAT).....	70
User Added Keypresses (AKP)	70
Transcription Average (STA)	74

CHAPTER 5: DATA BRIDGING	82
Background	82
Uninvestigated Interactions	83
Background	83
Identifying uninvestigated interactions	85
Additional Design Points	88
Background	88
Statistical criteria for selecting additional design points	92
Selecting Additional Design Points	93
Preliminary analysis	93
Development of a computer program	98
Selection procedures	101
Experiment 4	106
Results and Analysis	113
CHAPTER 6: INTEGRATED EMPIRICAL MODELS	116
Model Selection	121
Selection Procedures	124
Integrated Empirical Models	126
Interpretation of Empirical Models	129
Further Analysis	132
CHAPTER 7: OPTIMIZATION	135
Optimum Configurations	136
Optimum Configuration for TST Measure	136
Optimum Configuration for UAT Measure	140

Optimum Configuration for AKP Measure	143
Optimum Configuration for STA Measure	146
Analysis of Optimum Solutions	149
Summary of Optimum Solutions	150
CHAPTER 8: RULES AND CONSTRAINTS OF	
SEQUENTIAL RESEARCH STRATEGY	155
Initial Selection	156
Screening Study	156
Description	160
Experimental Designs	162
Identifying Uninvestigated Interactions	165
Conducting Sequential Experiments	166
Data Bridging	166
Comparability of sequential studies	166
Selecting additional design points	167
Discussion of constructing $ X'X $ matrix	169
Data collection at additional design points	170
Empirical Model Building	170
Administrative Requirements	171
Optimization	172
CHAPTER 9: CONCLUSIONS	174
Integrated Empirical Models of a Telephone	
Information System	174
Sequential Research Strategy	177
Future Research	178

REFERENCES	183
APPENDIX	193
Appendix A: Constant Factor Levels	193
Appendix B: Experimental Data Points and Dependent Measures	196
Appendix C: Reanalysis Results of Screening Study	203
Appendix D: Reanalysis Results of Previous Studies	208
Appendix E: Informed Consent Form	217
Appendix F: Pre-Test Questionnaire	221
Appendix G: Subject's Instructions	223
Appendix H: Targets and Information Messages	231
Appendix I: After-Test Questionnaire	234
Appendix J: Theoretical Background for Data Bridging	237
Appendix K: Computer Program List	248

LIST OF ILLUSTRATIONS

Figure 1. Iterative procedures of human computer interface design 4

Figure 2. Three stages of sequential research strategy ... 12

Figure 3. Procedures of dissertation research 16

Figure 4. Major component of a telephone information system 19

Figure 5. Alternative ways of constructing telephone information systems 20

Figure 6. 2x6 database structure 24

Figure 7. 8x2 database structure 26

Figure 8. Processes used to define experimental space for a screening study 32

Figure 9. Configuration of experimental apparatus 59

Figure 10. Significant main effects for total search time 72

Figure 11. Significant main effect for user added keypresses 77

Figure 12. Significant main effect for transcription average 79

Figure 13. Significant interaction for transcription average 80

**Figure 14. Some examples where it is impossible to
investigate a two-factor interaction with
four design points 84**

**Figure 15. Plot of total search time using input timeout
and background music level 151**

**Figure 16. Plot of transcription average scores using
speech rate and background music level 152**

Figure 17. Tools for conducting sequential experiments .. 179

LIST OF TABLES

Table 1.	Major Categories of Independent Variables	29
Table 2.	Independent Variables of Screening Study	33
Table 3.	ANOVA Results of Screening Study	40
Table 4.	Final Candidates for the Description Stage	41
Table 5.	Series of Experiments in the Description Stage	43
Table 6.	Factor Levels in Experiment 1	45
Table 7.	Factor Levels in Experiment 2	46
Table 8.	Factor Levels in Experiment 3	49
Table 9.	Defining Relationship and Design Structures of Experiment 3	52
Table 10.	Allasing Structures of Experiment 3	54
Table 11.	Treatment Conditions in Experiment 3	55
Table 12.	Sentence Structures of Product Information Messages	62
Table 13.	Experimental Procedure	63
Table 14.	ANOVA Summary Table for Total Search Time (TST)	71
Table 15.	ANOVA Summary Table for User Added Time (UAT)	73

Table 16. ANOVA Summary Table for User Added	
Keypresses (AKP)	76
Table 17. ANOVA Summary Table for Transcription	
Averages	78
Table 18. Results of Newman-Keuls Tests for ITxSEX	
Interaction	81
Table 19. Main Effects Investigated in Each Sequential	
Study	86
Table 20. Two-Factor Interactions Investigated in Each	
Sequential Study	87
Table 21. Uninvestigated Two-Factor Interactions During	
the Sequential Studies	89
Table 22. The Fourth Design Point for the Uninvestigated	
Interactions in the Telephone Information	
Study	91
Table 23. Significant Effects Resulting From the	
Sequential Studies	95
Table 24. Analysis Results of SEX Interaction Effects	97
Table 25. Reduced Set of Uninvestigated Two-Factor	
Interactions	99
Table 26. Maximum $X'X$ Value for Each Uninvestigated	
Interaction (First Design Point)	102
Table 27. Maximum $X'X$ Value for Each Uninvestigated	
Interaction (Second Design Point)	105

Table 28. Analysis of VIF and Eigen Value Ratio for TST (With Two Additional Design Points)	107
Table 29. Analysis of VIF and Eigen Value Ratio for UAT (With Two Additional Design Points)	108
Table 30. Analysis of VIF and Eigen Value Ratio for AKP (With Two Additional Design Points)	109
Table 31. Analysis of VIF and Eigen Value Ratio for STA (With Two Additional Design Points)	110
Table 32. Analysis of VIF (With Six Additional Design Points)	111
Table 33. Treatment Conditions of Additional Design Points	112
Table 34. Results of Partial-F tests for Uninvestigated Interactions	114
Table 35. Results of Regression Analysis for TST	117
Table 36. Results of Regression Analysis for UAT	118
Table 37. Results of Regression Analysis for AKP	119
Table 38. Results of Regression Analysis for STA	120
Table 39. Results of Model Selection for UAT	125
Table 40. Results of Regression Analysis for UAT Using a New Model	127
Table 41. Conversion of Variables From Coded Values To Natural Values	130
Table 42. Model Items Used for Developing Empirical Models	131

Table 43. Eight Equations of TST for Different Combinations of Discrete Factors	138
Table 44. Optimum Configurations for TST Measure	139
Table 45. Eight Equations of UAT for Different Combinations of Discrete Factors	141
Table 46. Optimum Configurations for UAT Measure	142
Table 47. Sixteen Equations of AKP for Different Combinations of Discrete Factors	144
Table 48. Optimum Configurations for AKP Measure	145
Table 49. Four Equations of STA for Different Combinations of Discrete Factors	147
Table 50. Optimum Configurations for STA Measure	148
Table 51. Summary of Optimum Solutions	154

GLOSSARY

- 2^k Factorial Design: An experimental design which includes k two-level factors.
- 2^{k-p} Fractional Factorial Design: A factorial design which uses a p-th fractional of a 2^k factorial design.
- ANOVA: Analysis of Variance.
- Average CI: The average width of confidence intervals of a regression model.
- Canonical Analysis: The reduction of the response surface to canonical form.
- CCD: Central Composite Design.
- Coded Value: An artificial number for expressing the real values of the levels of each factor.
- Cp Statistic: A statistical criterion to calculate prediction variance and bias of a regression model.
- Data Bridging: Statistical methods for integrating several sequential studies.
- Dependent Variables:
 - Total Search Time (TST): Average total search time to access an information message.
 - User Added Time (UAT): Average search time used to access a message in excess of the minimum search time required to complete an error-free search.
 - User Added Keypresses (AKP): Average number of keypresses used to access a message in excess of the minimum number of keypresses required to complete an error-free search.
 - Transcription Average (STA): Average number of correctly transcribed

keywords in an information message (four keywords scored per message).

- Design Point: An experimental treatment condition.
- Design Resolution: A statistical criterion for determining confounding structures of a fractional factorial design.
- DFFITS: Influence on the fitted value.
- Experimental Region: An experimental data space generated by independent variables.
- df: Degree of Freedom.
- Fractional Factorial Design: An experimental design using a fraction of treatment conditions of a full factorial design.
- Generating Design Point: An experimental treatment condition which includes only the factor levels specified in the fourth design points of uninvestigated interactions.
- Group Screening Design: An experimental design which uses a group of factors as a single factor.
- HCI: Human Computer Interface.
- Independent Variables:
 - Age of Subject (AGE): Age of the user.
 - Database Organization (DB): Length of menus and number of menus required to reach a message.
 - Feedback (FB): Augmented feedback to confirm user actions.
 - Input Timeout (IT): Time between the presentation of each menu item.
 - Background Music Level (MUS): Level of background music presented at the user's ear.

- Number of Targets (NT): Number of target items searched for in a single telephone call.
 - Sex of Subject (SEX): Sex of the user.
 - Speech Rate (SR): Rate used to present the synthesized speech.
 - Type of Voice (TV): Sex of synthesized speech.
 - Wallet Guide (WG): A visual diagram of the menu structure and menu items used to seek messages.
- Integrated Empirical Models: Empirical models based on several data sets.
 - MS: Mean Squares.
 - MSE: Mean Square Error.
 - Outliers: Observations which do not fit the trend set by the balance of the data.
 - PRESS: Prediction Sum of Squares.
 - Ridge Analysis: A statistical analysis method to find optimum solutions.
 - R-Student Statistic: A statistic for detecting outliers.
 - Sequential Experiments: A series of experiments which follow a common experimental procedure using a different set of independent variables.
 - SS: Sum of Squares.
 - Super Saturated Design: An experimental design which requires fewer experimental runs than the number of factors of interest.
 - Uninvestigated Interactions: Two-factor interactions uninvestigated during the sequential studies.
 - VIF: Variance Inflation Factor.
 - X matrix: The model matrix of an empirical model.

CHAPTER 1

INTRODUCTION

Design and Evaluation of Human-Computer Interfaces

Design Procedures

Design and evaluation of human-computer interfaces (HCIs) is a complex procedure. This procedure requires simultaneous consideration of many factors such as the user, task, hardware, software, and environmental factors. Traditional human factors research of HCIs has been restricted primarily to hardware and workplace layout considerations. Although these considerations are important for the design of HCIs, usable interfaces can be developed only when the complex relationships among the factors are considered.

Recently, human factors research has focused more on the user factor than the others in the design of an interface. This approach is based on the new concept, i.e., usability, that any system designed for people to use should be easy to learn and pleasant to use (Gould and Lewis, 1985). According to Shackel (1984), the working definition for the usability of a system (or an interface) is "the capability in human functional terms to be used easily (to a specified level of subjective assessment) and effectively (to a specified level of performance) by the specified range of users, given specified training and user

support, to fulfill the specified range of tasks, within the specified range of environmental scenarios.” The usability of a system is measured by both subjective and objective metrics such as learnability, throughput, flexibility, and attitude.

Based on the concept of usability, a variety of approaches have been proposed as a means of providing the procedures and principles of designing and evaluating user interfaces. These approaches are more user-oriented and involve iterative procedures during the development of an interface.

An approach proposed by Bennett (1984) and Shackel (1984) uses measurable goals for planning and controlling the development of an interface. This approach divides the development project into three distinct stages: (1) specifying user performance objectives as part of the product definition stage; (2) applying human factors principles, guidelines, and methods during the design and implementation stage; and (3) performing empirical testing of user performance to evaluate whether the product satisfies the user performance objectives. Gould and Lewis (1983; 1985) provide three critical concepts that need to be considered in developing an interface. These concepts include: (1) early focus on users and tasks; (2) empirical measurement of usage; and (3) iterative design. Using these three design principles, their approach consists of two stages: preliminary design stage and iterative testing and design. Another design procedure which improves the approach proposed by Bennett (1984) and Shackel (1984) includes the steps of defining usability through quantitative metrics, setting planned levels of usability, analyzing the impact of design solutions, incorporating user-defined feedback, and iterating the interface design until the planned levels of usability are achieved (Whiteside, Bennett,

and Holtzblatt, 1988). They emphasize importance of the natural context of actual use to collect richer information on the usability of an interface. The natural context includes work, time, social, and motivational context.

Williges, Williges, and Elkerton (1987) formalize the design procedures as an attempt to integrate these approaches to designing software interfaces. Their procedures consist of three iterative steps: (1) initial design; (2) formative evaluation; and (3) summative evaluation. Figure 1 shows the three iterative stages of their design procedures. Although the iterative design procedure is typical of accepted human factors design procedures, the specific research methods for evaluating and thereby providing feedbacks can vary according to the objectives of development and to the domain of interface design.

Research Methods

A variety of research methods have been proposed to design and evaluate human factors interfaces. For example, Ray and Ravizza (1986) characterize possible research approaches in HCI designs in terms of three dimensions: (1) naturalistic observations versus true experiments; (2) field research versus laboratory research; and (3) scientist as participant versus scientist as observer. There is therefore a rich space of possible research methods. Different points in this space have unique characteristics in applying the results of a study.

Williges (1981) provides a different point of view in classifying human factors research in complex systems. He summarizes three general approaches in conducting research which primarily deals with complex relationships among many independent variables. One approach is to make

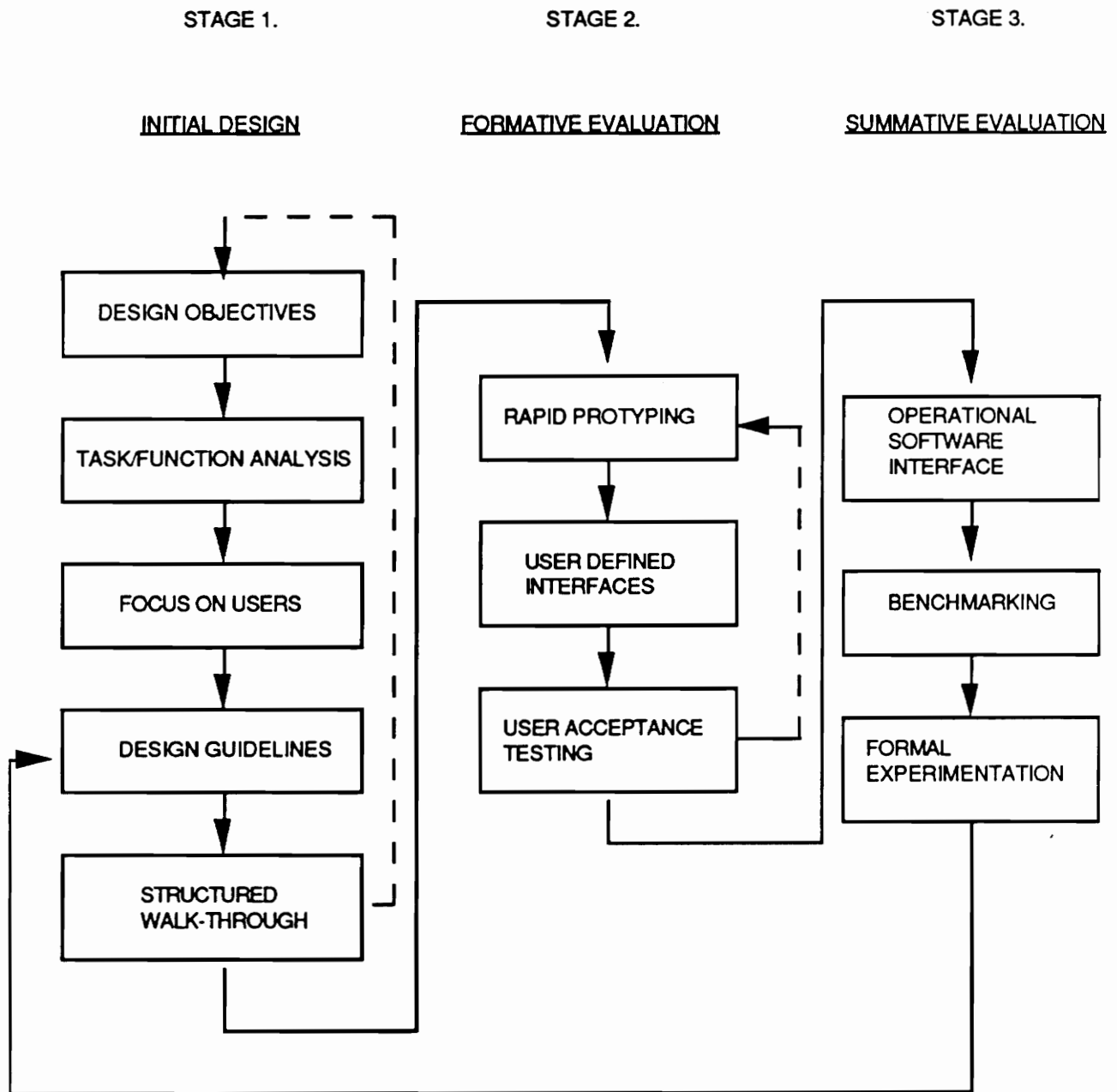


Figure 1. Iterative procedures of human computer interface design
(From Williges, Williges, and Elkerton, 1987).

direct observations of the operational system (or a prototype), which are major data collection techniques in usability engineering. A second approach is to use various analytical methods of modeling and fast-time computer simulation. The third approach is to use real-time simulation and manipulate various systems parameters according to experimental design methods.

Given the complexity of the design of HCI, typical research methods in HCI design are to make direct observations of the operational system (usability engineering) or to use laboratory experiments which abstract the real-world representations.

Approaches in usability engineering. The primary research tool of the usability engineering approach is to use direct observations of the operational system and contextual interviews with users. In general, this approach has a constructive influence on design and it helps the designer identify and propose solutions for users' problems. Further, an iterative design helps the designer manage system development since modification can be made step by step (Whiteside, Bennett, and Holtzblett, 1988). Another characteristic of this approach is the use of a small number of subjects during the testing. Since there is no attempt to generalize the results beyond the particular design in question, and since no attempt is usually made to achieve statistically significant results, it may not be necessary to employ a large number of subjects during the testing (Mills, 1986). Therefore, the results from the development of a specific design cannot be applied to the design of other interfaces even though standardized testing procedures and tasks are used during the testing. Thus, the results are usually task- and situation-specific.

Perhaps the major drawback of the usability engineering approach is the lack of control over independent variables. It may not be possible to determine an optimal configuration of a design since the approach does not allow the researcher to build functional relationships among the independent variables. Furthermore, the approach cannot guarantee that an independent factor has a significant influence on the design. When this approach is based solely on direct observations of only a few users, Landauer (1988) points out that often human ability to judge the determinants of human behavior is overestimated and variability in performance and preference are underestimated.

Laboratory experimentation. Formal laboratory experiments usually provide a high degree of control over independent variables. They provide less noisy and more objective data than direct observations and have been largely utilized for formative evaluations using mock-up or prototype systems (Carroll, Landauer, John, Whiteside, and Wolf, 1989). Usually a small number of independent variables is manipulated using typical experimental designs such as factorial designs. The major advantage of laboratory studies is that they can provide statistical procedures to identify significant factors and their interactions. Further, the results can be used to obtain the functional description of the effects of manipulated variables. Therefore, human performance in the system can be predicted and optimized based on the results of a study.

However, the results of laboratory studies have been questioned in terms of their inefficiency and lack of generalizability (Simon, 1987; Smith, 1987). One of the major reasons is the use of a small number of independent variables manipulated in an individual study. Often the results of these experiments are

of limited use in the design of real interfaces because many relevant factors have not been considered simultaneously in the experimental design used.

Problems of human factors research. In 1987, Smith made a critical evaluation and challenge to the human factors profession by stating: " Why, after all these years of research and investigation in the same fundamental areas, is there no progression in the level of knowledge? ... There has been no progress in the accumulation of general knowledge in the realm of human factors." The main reason, he states, is that most human factors research has not been designed for generalizability. That is, most human factors research has to be specific because the researchers (or designers) are often asked to provide definite answers and numbers for specific designs. The results of a specific study are therefore of limited use in the design of real systems due to the lack of generalizability. Chapanis (1988) defines a generalizable result as a finding which holds in situations other than the one in which it was observed. By definition, generalizability involves an extrapolation to conditions not identical to those at the time original observations were made.

The lack of generalizability of human factors research was predicted in the early 1970's. Simon (1973) pointed out that most human factors research uses such small experimental designs that the results are rarely related to human performance in the real world. This conclusion was based on the results of his review of 239 experiments published in Human Factors showing that the median number of factors studied in these experiments was two.

In addition to the small number of factors examined in these formal studies, most reports do not present the levels of factors held constant during the experiments. The failure to report this information may contribute to the lack

of generalizability since the researcher often needs to use the information in evaluating and designing interfaces.

Generalizable results should involve large data spaces with many factors and a defining relationship so that systematic predictions or extrapolations are feasible. The large number of factors makes data collection economy an important issue in conducting formal experiments. Human factors research based on traditional experimental designs becomes unreasonable due to resource and/or time constraints when studies include a large number of independent variables. For example, research in human-computer interaction usually involves many factors because interactions among many factors play an important role in the design of an interface (Landauer, 1988). However, due to the complexity of the human, task, hardware, software, and environmental factors, it is difficult to investigate the general relationships among the factors only with a single experiment or even with several unrelated experiments. As a result, the results of each study can be applied only to the specific design in question rather than to answering many questions about the consequences of design variations.

Empirical Models as Design and Evaluation Tools

Models are alternative ways of describing things as a means of finding out something about the real world (Chapanis and Van Cott, 1972). Generally a model is more abstract than the thing it describes. Due to the complexity of the real world, models can be used in the design and evaluation process since

they permit the researcher to reason more effectively about certain aspects of the thing modelled.

Although there are several criticisms of the model view (Thimbleby, 1990), and the goals of research and design as they relate to models are different (Williges, 1987), the use of models in HCI design can provide the designer with appropriate tools for efficient interface design. Various models for interface design tools can be divided into the broad categories of conceptual models, which deal primarily with representing cognitive processes, and quantitative models, which deal primarily with numerical representation of user performance (Williges, 1987). The conceptual models may provide descriptive properties of information processing activities based on characteristics of cognitive processes, structures, and strategies; whereas the quantitative models may predict user performance based on characteristics of tasks, users, and working environment.

Quantitative Models

One of the most useful approaches to building quantitative models is the use of polynomial regression. Polynomial regression is a useful procedure for prediction of human performance, variable screening, model description, and system parameter estimation (Myers, 1990). This procedure has been used to model human performance with various interfaces (Spine, Williges, and Maynard, 1984; Williges and Williges, 1982) and appears to be useful in assessing the effects and trade-offs of variables affecting HCIs (Williges, 1987). An important advantage of using polynomial regression as a means of building a quantitative model is that the results of several sequential studies can be combined into a complex regression equation.

To build an empirical model using a polynomial expression between human performance and its affecting variables, the researcher must determine what order of model should be used and identify what independent variables should be included in the model. The appropriate order of a model should be determined to describe adequately the complex relationships of independent variables. Usually a second-order polynomial approximation is adequate to account for most human performance (Williges, 1981). For example, the second-order polynomial regression equation using three independent variables would be:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_1X_2 + b_5X_1X_3 + b_6X_2X_3 + b_7X_1^2 + b_8X_2^2 + b_9X_3^2 \quad \text{Eq. (1)}$$

where Y is the predicted value of human performance, $b_0 - b_9$ are coefficients of model items, and $X_1 - X_3$ are independent variables.

Sequential Research Strategy as a Tool for Empirical Model

Building

Once the researcher has defined an appropriate empirical model, data must be collected to obtain estimates of the coefficients of the model. Although happenstance data can be used to solve the equation, they could cause undesirable problems such as inconsistencies, limited range of variables, semi-confounding of effects, nonsense correlations, and time dependent effects (Box, Hunter, and Hunter, 1978). All these problems could be avoided through the systematic use of experimental designs.

Due to the complexity of most HCI designs, a large-scale, factorial, experimental design would be required to handle the large number of relevant variables. These experimental designs, however, quickly become inefficient in

terms of the number of data points required. On the other hand, the use of small experiments cannot provide complete information for the design of real interfaces due to their lack of generalizability. Although it is a common practice in the design of an interface to use standardized benchmark tasks, performance measures, and experimental procedures, integration of the data sets from several individual studies needs to be conducted to obtain generalizable design implications. As an alternative, Williges and Williges (1989) proposed an efficient and systematic approach to considering multiple factors simultaneously without losing generalizability of the results while maintaining data collection economy.

The sequential research strategy is a set of methodological tools which has been developed to provide functional relationships among many factors in an efficient manner. This approach attempts to combine the data sets resulting from several experiments as a means of developing integrated empirical models. This approach consists of three major phases: selection, description, and optimization of independent variables (Figure 2).

Selection. In the selection of independent variables, all independent variables which could affect the human performance in a system are identified and justified. Since the number of independent variables, at this stage, is usually so large as to preclude the use of a full factorial design, an efficient screening study should be developed to reduce the number of independent variables.

Description. This stage requires a systematic procedure to divide the total set of independent variables into smaller subsets since the number of independent variables may be so large that the use of a full factorial design is

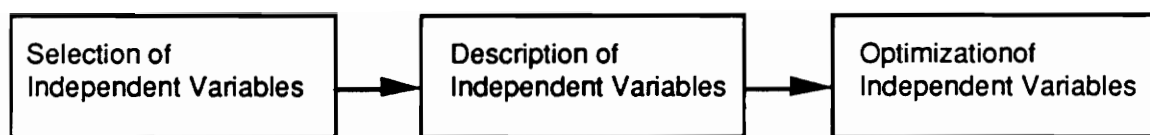


Figure 2. Three stages of sequential research strategy
(From Williges and Williges, 1989).

not efficient for building second-order empirical models. Based on each subset of independent variables, a series of small studies is conducted to provide the functional relationships among the independent variables within each subset. The results of each individual study are then combined into a single data set. Usually additional data collection is required at this stage since the combined data set does not provide necessary information to build complete second-order empirical models. Empirical models are developed using the integrated data set.

Optimization. Since the empirical models developed in the description stage provide the functional relationships among all important independent variables, the models can be analyzed to optimize the human performance. Optimization procedure should consider both continuous and discrete variables since both types of variables are usually involved in the design of human factors interfaces.

Research Objectives

The primary objective of this dissertation research is to provide a systematic way to combine several data sets into one integrated data set which will be used to build second-order empirical models in the design of a telephone information system. The secondary objective is to summarize the rules and constraints of the sequential research strategy based on the lessons learned from conducting the dissertation research.

Approaches to Achieving the Objectives

To demonstrate the use of the sequential research strategy in the design of HCIs, a telephone information system was selected. A study (Merkle, Beaudet, Williges, Herlong, and Williges, 1988) was conducted to select potential variables in the design of a telephone information system. Chapter 2 describes potential variables which can be used in the design of a telephone information system.

Primary Objective. The procedures for achieving the primary objective deal with the most efficient way to augment the existing data sets for the integration during the description stage. Based on the integrated data set, empirical models are built to describe the functional relationships among independent variables. The empirical models are then analyzed to provide optimum configurations of the interface.

To determine necessary conditions for combining data sets, reanalyses of previous studies were conducted. The previous studies include: a screening study (Beaudet and Williges, 1988), in which variables requiring investigation were selected from a large variable set; and a series of sequential studies (Martin, Williges, and Williges, 1990; Wu, 1989), in which the effects of a subset of independent variables were investigated. Chapter 3 presents the results of the reanalyses of Experiment 1 (Wu, 1989) and Experiment 2 (Martin, et al., 1990) along with the screening study (Beaudet and Williges, 1988). From the results of the reanalyses, a study (Experiment 3) was conducted to investigate three important independent variables which had not been examined in the previous studies. Detailed descriptions of the experimental design and methods used are presented in Chapter 4. Another experiment (Experiment 4)

presented in Chapter 5 was also conducted in order to bridge the data sets resulting from the previous studies. A computer program was developed to select the design points necessary for the Experiment 4. Using the integrated data set, Chapter 6 presents the procedures of developing integrated empirical models and Chapter 7 illustrates the analysis for obtaining the optimal configuration of a telephone information system.

Secondary Objective. To achieve the secondary objective, necessary conditions and techniques for conducting sequential experiments were examined. These include administrative requirements, alternative experimental designs, methodological considerations on conducting sequential experiments, and other necessary rules and decision criteria for bridging data sets and optimizing empirical models. A detailed description of the rules and constraints is presented in Chapter 8 of this report. This chapter describes general procedures and guidelines of applying the sequential research strategy to the design of other HCIs.

Figure 3 briefly illustrates the procedures to achieve the objectives. A detailed description of these procedures is presented in the following chapters.

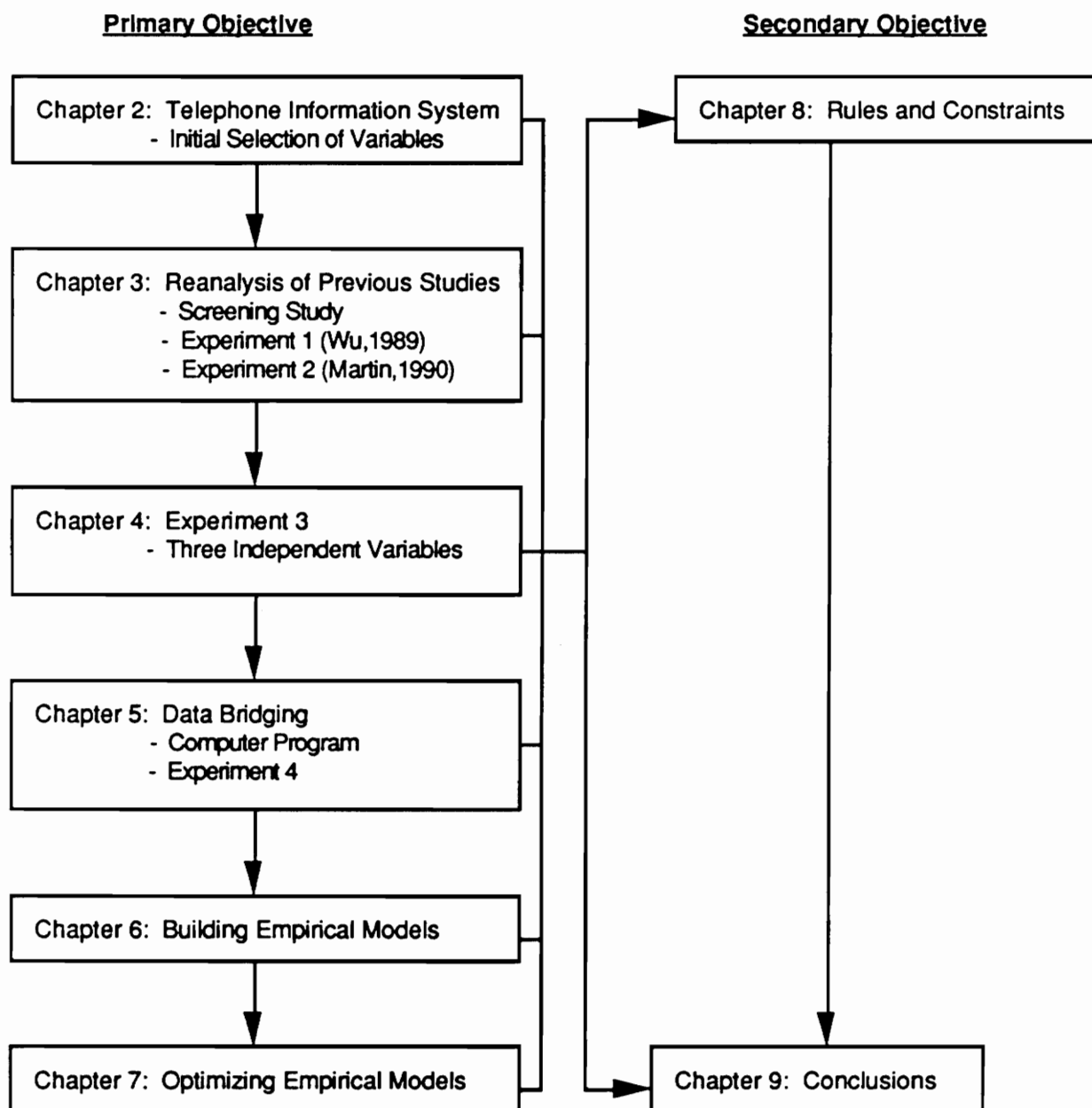


Figure 3. Procedures of dissertation research.

CHAPTER 2

TELEPHONE INFORMATION SYSTEMS

Background

Information systems enable the user to seek necessary information in a systematic manner. Telephone books, library catalogs, and department store catalogs, etc. are designed to provide information for a specific subject area. Major displays in the design of information systems can be divided into two general categories: visual displays and auditory displays. Visual displays are useful for presenting complex and parallel information, while auditory displays are useful for presenting simple and serial information (Sorkin, 1987). Visual displays have been a major tool for presenting information to the information seeker, since visual displays are popular and appropriate for conveying information at high rates.

Recent advancements of speech synthesis technology allow auditory displays to have a wider application area. Telephone information systems with synthesized speech can provide a variety of applications including home information services, remote office functions, database access, and information management for the visually impaired. In a telephone information system, the user can communicate with computers which manage databases and speech

synthesizers. Figure 4 illustrates the basic components of a telephone information system.

Since telephones are ubiquitous and connectable to computers, there are several circumstances where telephone information systems might be useful as a means of retrieving information in office applications (Berney, 1984; Gould and Boies, 1983). These applications involve primarily remote-information retrieval where visual displays are not readily available. In fact, synthesized speech has been used for a variety of applications including a menu-driven telephone inquiry service (Witten and Madams, 1977), voice-mail systems (Gould and Boies, 1984; Schmandt, 1985), an information service system (Podgorny, 1985), a tourist information system (Thomas, Rosson, and Chodorow, 1984). These applications use synthesized speech as system outputs. In other words, information from the system is delivered to the user using synthesized speech.

Halstead-Nussloch (1989) categorized several components of a telephone information system in terms of selecting communication media and dialogue designs. Figure 5 presents alternative ways of constructing telephone information systems. Communication media consist of user input and system output. User can input information either by telephone keypad or by voice. Most telephone information systems use the keypad to send control-input to the computer although there is ongoing research on the use of telephones as data entry devices (Detweiler, Schumacher and Gattuso, 1990; Marics, 1990). Eventually automatic speech recognition may also be incorporated with telephone information systems, and a variety of research is being conducted to

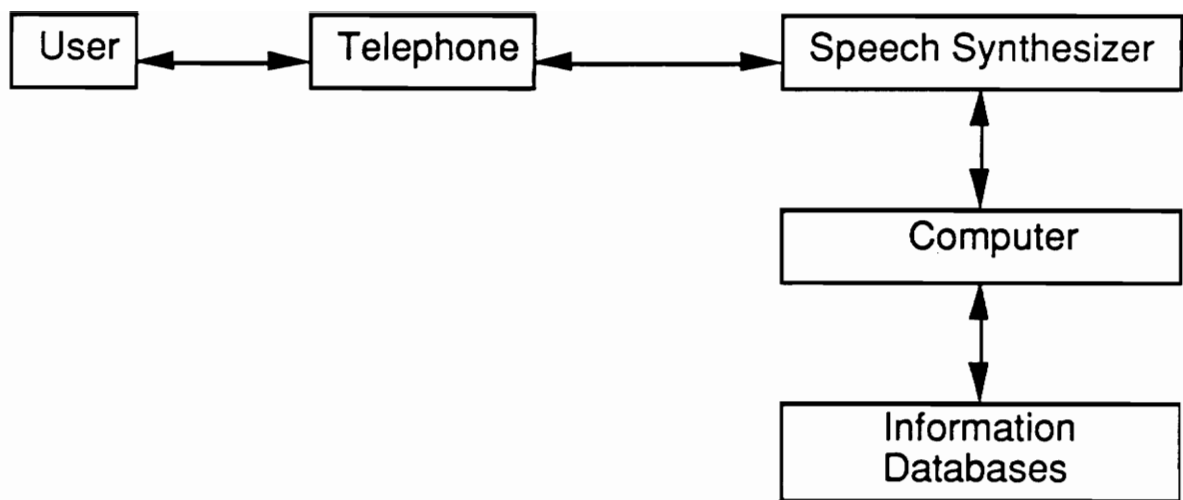


Figure 4. Major components of a telephone information system.

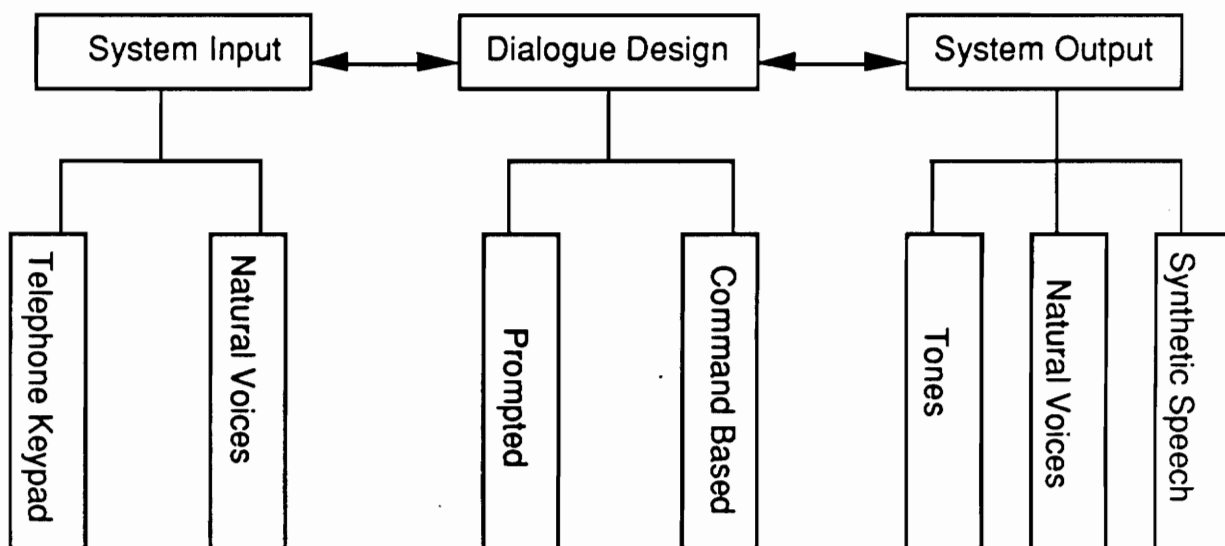


Figure 5. Alternative ways of constructing telephone information systems
(From Halstead-Nussloch, 1989).

improve the recognition accuracy of speech synthesizers (Simpson, McCauley, Roland, Ruth, and Williges, 1985).

Information output can be sent using either recorded, digitized voices or synthetic speech. Digitized speech generally provides better intelligibility than recorded voices (Halstead-Nussloch, 1989), and can be used in the applications where relatively a small amount of vocabularies are required. However, these techniques may be impractical for many large database inquiry and information systems since they require a large amount of computer storage (Lauretta, Rodman, and Antin, 1990; Sanders and McCormick, 1987). On the other hand, rule-based synthetic speech uses algorithms containing rules of pronunciation to convert text to speech in real time (Streeter, 1988). The use of rule-based synthetic speech provides flexibility in the form of unlimited vocabulary with far fewer storage requirements although the voice quality is usually not as good as that achieved by digitized speech (Sanders and McCormick, 1987).

Dialogue modes classified by Halstead-Nussloch (1989) are divided into two types: command based dialogue, which is activated by issuing commands; and prompted dialogue, which is activated by selecting an object and its related action. There are a variety of guidelines for the computer-based dialogues (Kiger, 1984; Williges and Ehrich, 1986; Williges, Schurick, Spine, and Hakkinen, 1986; Williges and Williges, 1984). These guidelines, however, may not be appropriate for the design of speech displays since dialogue designs involving speech display are inherently sequential and transient, while visual displays provide parallel and permanent information.

Although telephone information systems are already available, a few researchers have investigated individual factors that influence human performance with telephone information systems (Gould and Boies, 1984; Kidd, 1982; Podgorny, 1985; Schmandt, 1985; Witten and Madams; 1977). No systematic research has been used to develop empirical models of performance with telephone information systems. These models can be used to provide guidance on many issues including improving speech quality, designing database organizations, providing feedback for user control, selecting user input methods for error detection and recovery, and designing user aids (Martin, et al., 1990). These design issues can be solved by simultaneously considering multiple factors in terms of developing the functional relationships among them. In other words, systematic research should consider all important factors in developing empirical models of performance within telephone information systems.

Hypothetical Telephone Information System

To build empirical models for the design of telephone information systems, a hypothetical telephone information system was selected. This hypothetical system was designed to allow for the manipulation of many independent variables in a realistic test environment and for use of the selection, description, and optimization of sequential experimentation. The telephone information system uses synthesized speech in response to touch-tone telephone input to provide product information in the auditory database of a hypothetical department store.

Characteristics of the Hypothetical System

Major components of the telephone information system include a DECTalk (V2.0) speech synthesizer, a VAX 11/750 mainframe system, a Panasonic VA-8205 Easaphone, a stereo set, and auditory databases. The synthesizer is capable of converting text to speech in real time and adjusting a variety of speech parameters (e.g., speech rate, type of voice, pitch, etc.), and provides superior performance compared with other similar speech synthesizers (Greene, Manos, Pisoni, 1984; Greene, Logan, and Pisoni, 1986). The computer system connects the synthesizer to the auditory databases to send real time outputs to the user. The stereo set is designed to simulate a noisy real world environment by producing background music.

The auditory databases contain product items which can be found in a typical department store. These product items are organized using two different hierarchical menu structures. That is, the databases consist of two different structures in terms of menu breadth and menu depth: 2x6 structure and 8x2 structure. The 2x6 structure has two different product categories within each level (or menu) and has six hierarchical levels. This structure is presented in Figure 6. On the other hand, the 8x2 structure has eight product categories within each level (or menu) and has two hierarchical levels. This structure is illustrated in Figure 7. These database structures were selected to examine tradeoffs between the breadth and the depth of the hierarchy in the design of an auditory information system. At the bottom level of the hierarchy, product information messages are listed for each product item selected.

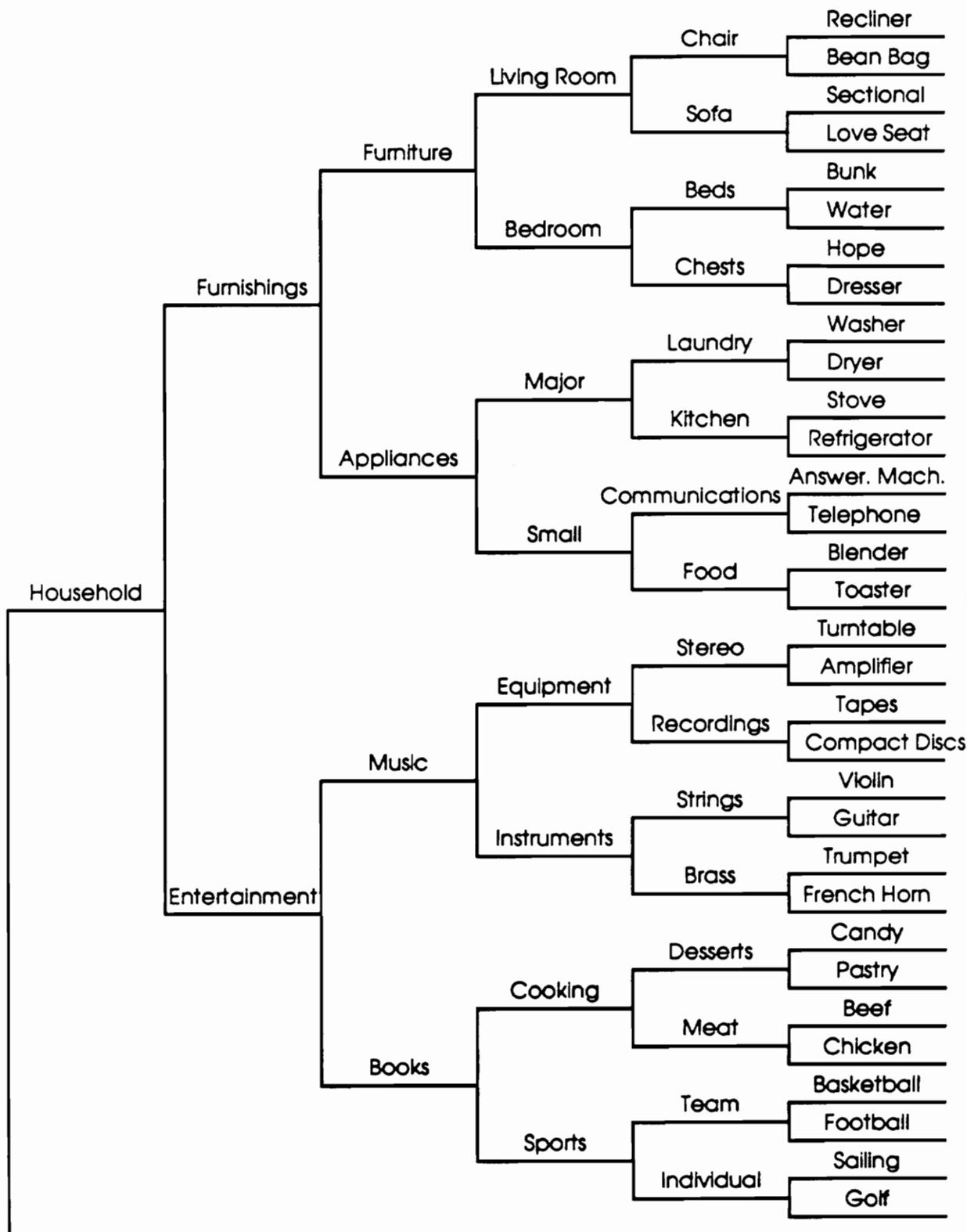


Figure 6. 2x6 database structure.

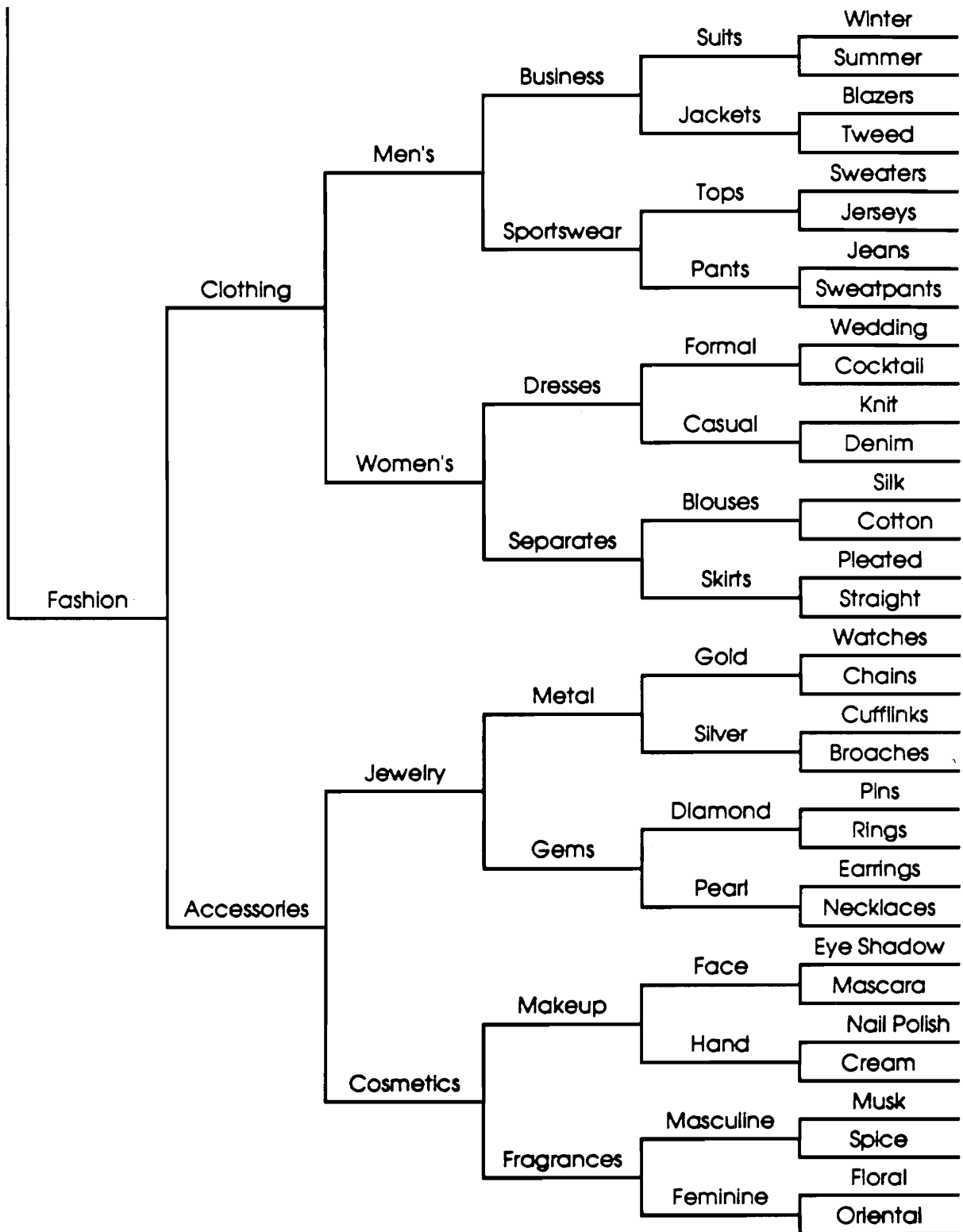


Figure 6. 2x6 database structure (Continued).

Furniture	Recliner Chairs
	Bean Bag Chairs
	Sectional Sofas
	Love Seat Sofas
	Bunk Beds
	Water Beds
	Hope Chests
	Dressers
	Washers
	Dryers
Appliances	Stoves
	Refrigerators
	Answering Machines
	Telephones
	Blenders
Music	Toasters
	Stereo Turntables
	Stereo Amplifiers
	Tape Recordings
	Compact Discs
	Violins
	Guitars
	Trumpets
Books	French Horns
	Candy Cookbooks
	Pastry Cookbooks
	Beef Cookbooks
	Chicken Cookbooks
	Basketball Books
	Football Books
	Sailing Books
	Golf Books

Figure 7. 8x2 database structure.

Men's Clothing	Winter Suits
	Summer Suits
	Blazers
	Tweed Jackets
	Sweaters
Women's Clothing	Jerseys
	Jeans
	Sweatpants
	Wedding Dresses
	Cocktail Dresses
Jewelry	Knit Dresses
	Denim Dresses
	Silk Blouses
	Cotton Blouses
	Pleated Skirts
Cosmetics	Straight Skirts
	Gold Watches
	Gold Chains
	Silver Cufflinks
	Silver Broaches
	Diamond Pins
	Diamond Rings
	Pearl Earrings
	Pearl Necklaces
	Eye Shadow
	Mascara
	Nail Polish
	Hand Cream
	Masculine Musk Fragrances
	Masculine Spice Fragrances
	Feminine Floral Fragrances
	Feminine Oriental Fragrances

Figure 7. 8x2 database structure (Continued).

Potential factors of interest. Using the configuration of a hypothetical telephone information system, a variety of potential factors which could affect performance at the interface were selected (Merkle, et al., 1988). These include speech quality, system dialogue design, user dialogue control, user characteristic, task characteristics, and environmental factors. Table 1 shows major categories of these potential factors. The procedures and methods that were used to select these potential factors are presented in the next section.

Initial Selection of Variables

In the first stage of a systematic research strategy, the researcher needs to define the total experimental space of independent variables which may affect performance in the system (Diamond, 1981). In complex systems, the experimental space usually includes a large number of potential factors such that a systematic approach needs to be developed in order to identify and justify a more refined experimental space which includes only the important factors. Many human factors researchers and statisticians proposed a variety of techniques for investigating a large data space in a systematic manner. Some examples include a sequential experimental plan as a means of building empirical models in chemical processes (Box, Hunter, and Hunter, 1978), and a research paradigm which systematically investigates complex systems (Williges and Mills, 1982). However, these techniques only present an approach to constructing an efficient experimental plan assuming that the researcher has already identified the list of independent variables. That is, these approaches do not provide a systematic tool for identifying the initial set of potential factors. To identify and justify all possible factors which could affect performance in a complex system, the researcher needs to gather information

TABLE 1

Major Categories of Independent Variables (From Williges and Williges, 1989)

<u>Speech Quality</u>	Error Recovery (4)
Speech Rate (4)*	User Assistance
Amplitude (1)	HELP System (5)
Harmonic Structure (14)	Embedded Training (1)
Prosodics (3)	Other User Aids (4)
Regional Accent (1)	<u>User Dialogue Control</u>
Exception Dictionary (1)	Speech Quality (2)
<u>System Dialogue Design</u>	Speech Pacing (5)
Speech Displays	Sequence of Events (2)
Vocabulary Design (4)	<u>User Characteristics</u>
Syntactical Structure (4)	Experience (4)
Semantic Structure (2)	Demographics (5)
Information Coding (1)	<u>Task Characteristics</u>
Speech Rate (1)	Response Time (1)
Keypad Input	Database Structure (4)
Input Disambiguation (2)	Task Complexity (6)
Input Coding Level (1)	Competing Task (1)
Menu Design (3)	<u>Environmental Factors</u>
Command Style (4)	Competing Speech (1)
Error Handling	Noise (1)
Error Detection (1)	Background Music (1)

* Number in parentheses indicates the number of factors.

regarding the performance of the system. Simon (1977) suggests five general sources of information for the initial identification of potential factors. These include: (1) literature review; (2) interview; (3) direct observation; (4) personal experience; and (5) data collection.

Initial selection of independent variables. As shown in Table 1, a list of potential factors was identified by brainstorming and then categorized by a variety of techniques including literature review, feasibility analysis, and relevance analysis (Merkle, et al., 1988).

The total number of potential variables was 95, which was large enough to preclude the simultaneous investigation of all factors. These potential factors, however, can be reduced to find a smaller experimental space in which only important variables are included. That is, the potential factors should be evaluated and validated through systematic screening studies since one can assume that a smaller number of factors will have significant influence on the performance.

Subjective ratings based on a prototype system were made to evaluate and validate the effects of the independent variables on the speed and accuracy of transmitting information, and the acceptability of the information system. The results of the validation study were integrated with a literature review to establish a set of decision rules regarding the reduction of the independent variables (Merkle, et al., 1988). When the effect of a variable was clearly known in the literature, the variable was set as a constant for further studies. When the effect of a variable was in conflict between the two sources, or the ratings had no clear opinion, the variable could be eliminated depending

upon the strength of the information from each source. Figure 8 illustrates the processes of refining the experimental space described in this study.

Screening of initial variables. Using these decision rules, a subset of 16 candidates for a screening study was selected from the original set of 95 independent variables. These variables include speech display factors (type of voice and speech rate), database (database organization and number of target items), system control rate (input timeout and system response time), command features (pause/resume speech, repeat speech, and spell-out speech), user aids (wallet guide, menu selection feedback, database selection feedback, and command selection feedback), environment (background music), and user demographics (age of subject and sex of subject). Table 2 shows these factors with their corresponding factor levels used in the screening study (Beaudet, 1988). The other variables are held constant during the subsequent experiments since these are found to have no significant effects on the design of a telephone information system. The values of these factors are presented in Appendix A.

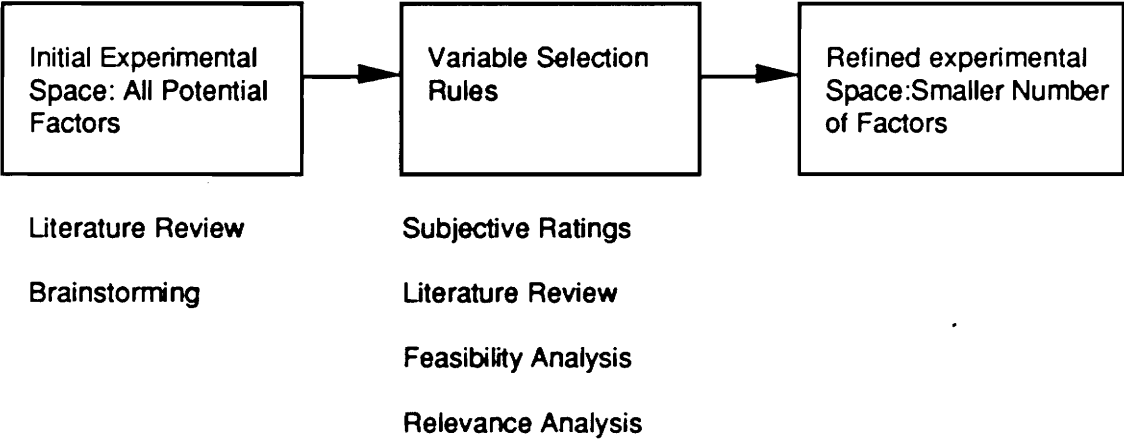


Figure 8. Processes used to define experimental space for a screening study (From Williges and Williges, 1989).

TABLE 2

Independent Variables of Screening Study (From Beaudet and Williges, 1988)

Variable	Level
<u>Speech Display</u>	
Type of Voice	Paul, Betty
Speech Rate	180, 240 wpm
<u>Database</u>	
Database Organization (b x d)	8 x 2, 2 x 6
Number of Target Items	One, Two
<u>System Control Rate</u>	
Input Timeout	2, 4 sec.
System Response Time	0, 4 sec.
<u>Command Features</u>	
Pause/resume Speech	Available, Not Available
Repeat Speech	Available, Not Available
Spell-out Speech	Available, Not Available
<u>User Aids</u>	
Wallet Guide	Available, Not Available
Menu Selection Feedback	Available, Not Available
Database Selection Feedback	Available, Not Available
Command Selection Feedback	Available, Not Available
<u>Environment</u>	
Background Music	Office, Mall Levels
<u>User Demographics</u>	
Age of Subject	College (18-30), Middle (45-60)
Sex of Subject	Male, Female

CHAPTER 3

REANALYSIS OF PREVIOUS STUDIES

Background

Integration of several data sets was not well planned when the previous studies (Beaudet and Williges, 1988; Martin, et al., 1990; and Wu, 1989) were conducted. In other words, a variety of constraints and requirements for data bridging were not established when the previous experiments were conducted. After the completion of Experiment 2 (Martin, et al., 1990), several analyses were conducted to determine necessary and sufficient conditions for integrating several data sets resulting from a series of sequential studies.

First, a feasibility study should be performed to determine whether the integration of data sets across studies is feasible. Second, the experimenter needs to identify what independent variables should be included in an integrated data set. Third, it is necessary to define the experimental space generated by independent variables as a means of determining appropriate design points for data bridging. Finally, the experimenter needs to determine whether further experimentation is required to integrate several data sets.

Feasibility Analysis

Integration of several data sets requires certain constraints. First, each data set must contain information on the factor levels of all factors which have

been either manipulated or held constant during the study. Second, each study must use the same experimental tasks and procedures to avoid data variations resulting from the use of different tasks and procedures. Since the individual studies during the description stage (Wu, 1989; Martin, et al., 1990) reported the levels of all factors and used the same experimental tasks and procedures, it is possible to integrate the data sets resulting from the studies. Further analyses showed that additional design points need to be added to the combined data set to estimate some uninvestigated interaction effects. Theoretical background and statistical criteria for selecting additional design points are presented in Chapter 5.

Experimental Space

Independent variables. The independent variables which will be included in the integrated data set should be carefully determined to describe the functional relationships explaining the performance of a system. Screening studies and a series of sequential studies can be used to determine what independent variables should be included in the integrated data set. In the study of the telephone information system, ten independent variables were found to be important to describe the performance at the interface. These variables resulted from the screening study (Beaudet, 1988).

Experimental space. The experimental space generated by the independent variables was defined as a spherical experimental region (Wu, 1989). This space was generated by four continuous variables (speech rate, input timeout, background music level, and age of subject). The information on the type of experimental region was used to determine the factor levels for the Experiment 3 and Experiment 4. A detailed description of the independent

variables and their corresponding factor levels is presented in the following section.

Analysis Procedures

Redefinition of Dependent Measures

To measure the human performance of the telephone information system, several dependent measures were defined (Beaudet, 1988). Among the dependent measures defined by Beaudet (1988), the following four dependent measures were redefined since some dependent measures were not used frequently (e.g., pause/resume selected total, repeat keyword selected total, and spell-out keyword selected total) and some measures were defined incorrectly (e.g., search time ratio and search efficiency ratio). Redefined dependent measures include total search time (TST), user added time (UAT), user added keypresses (AKP), and transcription average scores (STA).

Total search time is a score of how long it takes the subject to find a target. The total search time consists of two distinct parts: system imposed time and user imposed time. System time requirement is the minimum amount of time necessary for the system to speak the menu items. User imposed time is a score of how long it takes the subject to determine and select correct menu items. User imposed time consists of minimum target search time and user added time. The minimum target search time is determined by adding system time requirements plus 0.57 seconds for each menu level selection. The selection time of 0.57 is taken from the American Institutes for Research Data

Store (Munger, Smith, and Payne, 1962) for an ideal user pressing a pushbutton when cued and has a human reliability of 0.999 (Beaudet, 1988). The time value represents the average amount of time required for an individual to press a single pushbutton when a stimulus is presented. User added time is then calculated by subtracting minimum target search time from total search time. User added keypresses are calculated by subtracting minimum keypresses from total keypresses used to find a target. Minimum keypresses are a score of how many keypresses are necessary for an ideal user to find a target. The minimum keypresses for a 2x6 database are 7 (see Figure 6) while the minimum keypresses for a 8x2 database are 3 (see Figure 7). Only strict message transcription average (STA) is selected as the dependent measure for the message transcription task since the synonym message transcription average was highly correlated with the STA (Beaudet, 1988).

In summary, the new dependent measures which were used to reanalyze the previous studies include: total search time (TST), user added time (UAT), and added keypresses (AKP) for the target search task, and message transcription average (STA) for the message transcription task. Appendix B lists the values of these dependent measures along with the experimental design points of the previous studies.

Analysis Methods

Analysis of variance (ANOVA) methods were used to reanalyze the results of the previous studies with the redefined dependent measures. The results of the reanalysis were used to determine; (1) important independent variables resulting from the screening study (Beaudet, 1988); (2) an additional

study (Experiment 3) including three uninvestigated independent variables; and (3) significant effects for building integrated empirical models.

Results of Reanalysis

Screening Study

Following the initial selection of independent variables, efficient experimental designs are needed to screen the selected independent variables before committing to extensive data collection required to describe the relationships among the independent variables (Williges and Williges, 1989). Several experimental design alternatives are available to reduce the amount of data collection required in a screening study. Regardless of the design alternatives, at least Resolution III designs are recommended to keep all main effects of factors unconfounded (Myers, 1976).

In the telephone information study, a Hadamard matrix design (Resolution IV) described by Diamond (1981) was used to test efficiently the main effects of the 16 variables (Beaudet and Williges, 1988). The main effects were analyzed using only 32 data points in contrast to the 65536 design points required in a 2^{16} full factorial design. A lax criterion rather than the usual low alpha level (i.e., $\alpha = 0.05$) was used to prevent variables from being excluded too early in the study.

Based upon the new dependent measures, data from the screening study were reanalyzed by the ANOVA technique for four performance measures (total search time (TST), user added time (UAT), user added keypresses (AKP), and transcription average (STA)). Using the lax criterion ($\alpha=0.20$), important

independent variables were selected out of the 16 variables. These independent variables were found to be significant in at least one ANOVA test. Table 3 presents the results of the screening study. The results of each ANOVA for the four dependent measures are presented in Appendix C.

As shown in Table 3, five variables were found to be significant in at least two ANOVA tests; while eight variables were found to be significant in one ANOVA test. Before starting more detailed investigation of each candidate variable in Table 3, some changes were made due to unreasonable results or practical limitations. These changes and their corresponding reasons are listed as follows.

- (1) Repeat Keyword and Spell-out Keyword factors were dropped from the candidate list because these features of the telephone information system have rarely been used during search tasks.
- (2) Menu Feedback factor was combined with Selection Feedback factor, because they have the same characteristics in nature.

Based on these changes, ten variables (Table 4) were selected as final candidates for further investigation in the description stage.

Experiments for Description of Independent Variables

Following the screening study, a functional description of the relationships among the selected independent variables needs to be developed to predict the performance of the system. Since the number of the variables is so large as to preclude the use of complete factorial and multilevel experiments, a series of small experiments must be conducted in an

TABLE 3
ANOVA Results of Screening Study

Variable	Comment
Database Organization	Significant in at least two ANOVAs
Age of Subject	
Wallet Guide	
Selection Feedback	
Input Timeout	
Type of Voice (STA*)	Significant in one ANOVA
Repeat Keyword (STA)	
Sex of Subject (STA)	
Spell-out Keyword (STA)	
Menu Feedback (STA)	
Background Music (AKP)	
Number of Targets (AKP)	
Speech Rate (STA)	

*STA: Transcription Average
 AKP: User Added Keypresses

TABLE 4
Final Candidates for the Description Stage

Speech Rate	Input Timeout
Background Music	Age of Subject
Sex of Subject	Type of Voice
Wallet Guide	Database Organization
Feedback	Number of Targets

economical and strategic manner (Williges and Williges, 1989). Another reason to conduct a series of experiments rather than a single large experiment is that the set of the selected variables contains two different types of characteristics ,i.e., continuous variables and discrete variables. It is difficult to design an experiment, which contains both types of variables, for investigating linear effects and nonlinear effects in an efficient manner. Therefore, a systematic approach should be made to break down a complete set of variables into smaller sets on which an efficient estimation of both types of variables is achievable.

In the telephone information study, two criteria were used to divide the ten independent variables into smaller subsets: (1) selection of system parameters; and (2) the type of independent variables (continuous and discrete). First, speech rate and input timeout factors were selected as system parameters since they were considered important factors in describing the functional relationships and seemed to interact with other variables. Second, the variables were categorized into two groups: (1) continuous variables (speech rate, input timeout, age of subject, and background music); and (2) discrete variables (database organization, feedback, wallet guide, number of targets, type of voice, and sex of subject). The set of discrete variables was further divided into two groups; Group 1: wallet guide, database organization, and feedback; and Group 2: number of targets, type of voice, and sex of subject. The system parameters were included in every subset of variables. Based on the subsets of independent variables, Table 5 shows the series of experiments and their corresponding independent variables manipulated in

TABLE 5
Series of Experiments in the Description Stage

Experiment	Manipulated Variable	Constant Variable
Experiment 1 (Wu, 1989)	SR, IT, AGE, MUS	WG, DB, FB, SEX, NT, TV
Experiment 2 (Martin, 1990)	SR, IT, DB, FB, WG	MUS, SEX, AGE, TV, NT
Experiment 3	SR, IT, NT, SEX, TV	MUS, AGE, WG, DB, FB
* SR: Speech Rate IT: Input Timeout MUS: Background Music AGE: Age of Subject WG: Wallet Guide		DB: Database Organization FB: Feedback TV: Type of Voice NT: Number of Targets SEX: Sex of Subject

each experiment. The other variables in each experiment were held constant during the experiment.

Experiment 1. The set of continuous variables was examined in the first sequential experiment (Wu, 1989). Factors investigated were speech rate (SR), input timeout (IT), age of subject (AGE), and background music (MUS). Since the experimenter assumed that third- and higher-order interactions among the variables were negligible, a second-order central composite design ($\alpha = 1.414$) was used to evaluate four main effects, six two-factor interactions orthogonally. In addition, pure quadratic effects of the four variables were investigated. This design required five levels of each factor. The number of experimental design points for the factorial portion, axial portion, and center point were 16, 8, and 1, respectively. Second-order empirical models were developed using polynomial regression for the four performance measures. Table 6 shows the levels of the manipulated variables and constant values of the other variables during the first sequential experiment. Subsequent analyses of variance showed that SR, MUS, and AGE were found to be significant ($p < 0.05$). Only one two-factor interaction (IT x MUS) and one pure quadratic effect (MUS²) were found to be significant ($p < 0.05$).

Experiment 2. Five discrete variables were investigated in the second sequential experiment (Martin, et al., 1990). Factors investigated were speech rate (SR), input timeout (IT), wallet guide (WG), database organization (DB), and feedback (FB). A 2^{5-1} fractional factorial design (Resolution V) was used to investigate orthogonally all main effects and two-factor interactions since third- and higher-order interactions were assumed to be negligible. Table 7 shows

TABLE 6

Factor Levels in Experiment 1

Factor	Level
Manipulated	
SR*	120, 138, 180, 222, 240 wpm
IT	2, 3, 6, 9, 10 seconds
AGE	15, 22, 38, 54, 60 years
MUS	36.1, 40.5, 50.8, 61.1, 65.4 dBA
Constant	
WG	Not Available
DB	8 x 2
FB	Not Available
TV	Paul
NT	One
SEX	Uncontrolled
* SR : Speech Rate	WG : Wallet Guide
IT : Input Timeout	DB : Database Organization
MUS : Background Music Level	FB : Feedback
AGE : Age of Subject	NT : Number of Target Items
TV : Type of Voice	SEX : Sex of Subject

TABLE 7
Factor Levels in Experiment 2

Factor	Level
Manipulated	
SR*	120, 240 wpm
IT	2, 10 seconds
WG	Available, Not Available
DB	8 x 2, 2 x 6
FB	Available, Not Available
Constant	
AGE	College Age
MUS	Office (36.1 dBA)
TV	Paul
NT	One
SEX	Uncontrolled
* SR : Speech Rate	WG : Wallet Guide
IT : Input Timeout	DB : Database Organization
MUS : Background Music Level	FB : Feedback
AGE : Age of Subject	NT : Number of Target Items
TV : Type of Voice	SEX : Sex of Subject

the levels of the manipulated variables and constant values of the other variables in this experiment.

As shown in Table 7, the factor levels of the system parameters were set at the α value in the first sequential experiment. These values should have been set at +1 and -1 coded values since interaction information comes from the factorial part of a CCD. In other words, since the experimental region defined in the first experiment was spherical, the factors used in the first sequential experiment should have been set at +1 and -1 coded values in the subsequent experiments where only main effects and interaction effects were examined. Although the factor levels used in the second experiment can be used to develop integrated empirical models, it would have provided more detailed information with +1 and -1 coded values.

Analyses of variance showed that all main effects except FB were significant ($p < 0.05$). It is interesting to note that significant interactions in this study were the interactions among DB, IT, and WG ($p < 0.05$). That is, only three interactions (DB x IT, DB x WG, and IT x WG) were found to be significant. Appendix D presents the results of ANOVAs for Experiment 1 and 2.

Experiment 3. As shown in Table 5, two sequential experiments (Martin, et al., 1990; Wu, 1989) have been conducted during the description stage. Although the SEX variables were implicitly included in these two experiments, it was not manipulated to investigate its effect unconfounded with other main effects. In other words, it is impossible to evaluate the main effect of SEX and its interactions orthogonally since the variable was confounded with other manipulated variables in each experiment. Further, the other two independent variables (TV and NT) was not examined in the first two sequential

experiments. Therefore, another experiment (Experiment 3) was conducted to complete the functional description of the independent variables resulting from the screening study. Table 8 shows the levels of each manipulated variable and constant values of the other variables used in Experiment 3. A detailed description of experimental method and results of the Experiment 3 is presented in Chapter 4.

TABLE 8
Factor Levels in Experiment 3

Factor	Level
Manipulated	
SR*	138, 222 wpm
IT	3, 9 seconds
NT	One, Two
SEX	Male, Female
TV	Paul, Betty
Constant	
MUS	Office (36.1 dBA)
AGE	College Age
WG	Not Available
DB	8 x 2
FB	Not Available
* SR : Speech Rate	WG : Wallet Guide
IT : Input Timeout	DB : Database Organization
MUS : Background Music Level	FB : Feedback
AGE : Age of Subject	NT : Number of Target Items
TV : Type of Voice	SEX : Sex of Subject

CHAPTER 4

EXPERIMENT 3

Method

This experiment investigated the effects of five independent variables on human performance in the design of a telephone information system. Factors manipulated in this experiment were speech rate (SR), input timeout (IT), number of targets (NT), type of voice (TV), and sex of subject (SEX). SR and IT factors were included in this experiment since they were selected as system parameters. In other words, the interaction effects between the system parameters and the other factors (i.e., NT, SEX, and TV) were examined unconfounded with the other two-way interaction effects, although the main effects of SR and IT were manipulated in the previous experiments (Martin, et al., 1990; Wu, 1989).

This experiment was the last investigation of a subset of important independent variables resulting from the screening study. Experimental methods and procedures were the same as those used in the previous experiments to avoid data variations resulting from the use of different methods and procedures. The results of this study formed a component of an integrated data set which included all individual data sets resulting from the previous studies.

Experimental Design

All independent variables had only two levels each since the quadratic effects for the continuous variables (i.e., SR and IT) were examined in the previous study (Wu, 1989). A 2^{5-1} fractional factorial design (Resolution V) was used since this study investigated only the main effects and two-way interactions. That is, it was not necessary to use a 2^5 full factorial design to develop second-order empirical models since only main effects and two-way interactions would be included in second-order models for discrete variables. This experimental design was the same design used in Experiment 2. A between-subjects design was selected because one of the factors studied in this experiment was sex of subject and hence a within-subject design was not appropriate. Specific levels of each variable are presented in Table 8. As shown in Table 8, the levels of continuous variables (i.e., SR and IT) were set at the values corresponding to +1 and -1 (in coded values) of the CCD used in the first sequential experiment, because the experimental region of interest specified in the CCD was spherical. The levels of continuous variables were determined based on the fact that information on interaction effects is derived from the factorial portion of the CCD.

Table 9 shows the defining relationship used to select a one-half replicate of a 2^5 factorial design. As shown in Table 9, this experiment had 16 treatment conditions. In this design, higher-order interaction effects (i.e., three-way, and four-way) were confounded with main effects and two-way interaction effects. That is, all main effects were confounded with four-way interaction effects, and two-way interaction effects were confounded with three-way interaction effects. One five-way interaction effect was completely lost in this

TABLE 9

Defining Relationship and Design Structures of Experiment 3

 Defining Relationship $I = A + B + C + D + E^* = 0,1 \pmod{2}$

$I = 0$	$I = 1^{**}$
(1)	a
ab	b
ac	c
ad	d
ae	e
bc	abc
bd	abd
be	abe
cd	acd
ce	ace
de	ade
abcd	bcd
abce	bce
acde	bde
abde	cde
bcde	abcde

* $A=SR$, $B=IT$, $C=NT$, $D=SEX$, and $E=TV$

** $I = 1$ defining relationship was used in Experiment 3.

design. Table 10 shows the aliasing structures of this design. Specifically, the main effect of SR was confounded with the four-way interaction effect among IT, NT, SEX, and TV; and the two-way interaction effect between SR and IT was confounded with the three-way interaction effect among NT, SEX, and TV. Based upon the one-half fractional replicate selected, the levels of each factor were constructed by assigning +1 and -1 coded values and these coded values were converted into real treatment values for the experiment. Table 11 shows the 16 treatment conditions with the coded values and their corresponding real factor values. The other six factors were held constant during this experiment. The number of replications of each treatment condition was determined based upon the results of a power analysis (Martin, et al., 1990). With three replications per treatment conditions, the power (i.e., $1-\beta$) was approximately 0.88 when α equals 0.01. Each experimental run was randomly selected during the experiment.

Subjects

This study employed the use of 48 native, English speaking students at Virginia Polytechnic Institute and State University. Twenty four college age males and twenty four college age females participated in this experiment since sex of subject was a manipulated variable. Subjects completed an informed consent form (see Appendix E) and a pretest demographics questionnaire (see Appendix F).

Subjects were also given a hearing test with a Belltone 109 audiometer to make sure that they had normal hearing levels in both ears. During the hearing tests, they were given three, one-second long pulsed tones at 26 dBA for 750, 1000, 2000, and 4000 Hz. Tones were presented to the subject's right

TABLE 10

Aliasing Structures of Experiment 3

<u>Main Effects</u>	<u>Aliases</u>
A*	BCDE
B	ACDE
C	ABDE
D	ABCE
E	ABCD
<u>Two-Way Interactions</u>	<u>Aliases</u>
AB	CDE
AC	BDE
AD	BCE
AE	BCD
BC	ADE
BD	ACE
BE	ABD
CD	ABE
CE	ABD
DE	ABC
* A=SR, B=IT, C=NT, D=SEX, and E=TV	

TABLE 11

Treatment Conditions of Experiment 3

Treatment	<u>Coded Value</u>					<u>Real Value</u>				
	<u>SR</u>	<u>IT</u>	<u>NT</u>	<u>SEX</u>	<u>TV</u>	<u>SR</u>	<u>IT</u>	<u>NT</u>	<u>SEX</u>	<u>TV</u>
1	-1	-1	-1	-1	+1	138	3	1	F	P
2	-1	-1	-1	+1	-1	138	3	1	M	B
3	-1	-1	+1	-1	-1	138	3	2	F	B
4	-1	+1	-1	-1	-1	138	9	1	F	B
5	+1	-1	-1	-1	-1	222	3	1	F	B
6	-1	-1	+1	+1	+1	138	3	2	M	P
7	-1	+1	-1	+1	+1	138	9	1	M	P
8	-1	+1	+1	-1	+1	138	9	2	F	P
9	-1	+1	+1	+1	-1	138	9	2	M	B
10	+1	-1	-1	+1	+1	222	3	1	M	P
11	+1	-1	+1	-1	+1	222	3	2	F	P
12	+1	-1	+1	+1	-1	222	3	2	M	B
13	+1	+1	-1	-1	+1	222	9	1	F	P
14	+1	+1	-1	+1	-1	222	9	1	M	B
15	+1	+1	+1	-1	-1	222	9	2	F	B
16	+1	+1	+1	+1	+1	222	9	2	M	P

* SR=speech rate (words per minute) IT=input timeout (seconds)
 NT=number of targets (1 or 2) TV=type of voice (Paul, Betty)
 SEX=sex of subject (Male, Female)

Constants:

MUS=background music (36.1 dBA) AGE=age of subject (18-30 yrs)
 DB=database organization (8x2) FB=feedback (not available)
 WG=wallet guide (not available)

or left ear and they were required to raise their hands to indicate they heard the tones. If a subject were not able to hear the tones at a given frequency, the tones at that frequency and decibel level were repeated two more times. During retesting to pass the hearing test, subjects were required to indicate they heard correctly the tones both times. Any subject failing to pass the hearing test was not allowed to participate in the experiment.

Experimental Tasks

Two different tasks were performed during the experiment. Search tasks were performed to find information about product item(s) presented on a visual display by navigating an auditory database using a telephone keypad. After each search task, subjects were required to transcribe the information message about the product item. A detailed description of these tasks is presented in the experimental procedures.

Apparatus

Speaker telephone. A touch-tone speaker telephone (Panasonic Easaphone VA-8205) was used so that subjects did not have to hold a handset to their ear during the experiment. The telephone volume was set at approximately 64 dBA measured near subjects' ear. This constant volume level was used in all the previous studies and was not allowed to be altered to avoid the effects of the use of different volume levels.

This telephone had the standard layout of telephone keypads with a button for activating the speaker and several keys for storing telephone numbers and other functions. A specific telephone number was assigned on a key located on the upper right corner of the telephone. The telephone was placed at the front-right side of the subject on a terminal table. The keypads that

were used during the experiment include: the speaker button, the dial memory button, and four function keys for navigating the auditory database. These functions are explained in the experimental procedure section.

Speech synthesizer. A DECTalk (version 2.0) was used in this study. It is a rule-based speech synthesizer manufactured by Digital Equipment Corporation. The voices used in this study were Perfect Paul and Beautiful Betty with all the parameters at default settings. During an instruction session, these voices were used to present experimental procedures and the functions of several command keys on the telephone (see Appendix G). During the experimental sessions, these voices were used to provide product items and product information messages.

Speech parameters of the synthesizer such as pitch, regional accents, stress coding, etc. were not manipulated in this experiment. Product names and information messages were entered into the database as they are normally spelled and spoken; the only exception was that compound words were entered with hyphens at the appropriate location to avoid possible mispronunciation (e.g., foot-ball, sweat-pants).

Computer system. A VAX 11/750 computer system was used to present experimental tasks and collect dependent measures in real-time. This experiment had real-time priority over any other activities of the computer. The computer recorded all transactions with corresponding time information between the user and the system for the search and transcription tasks.

Two DEC VT220 visual display terminal were used. The terminal in front of the subject was used to present the names of targets and cue the subject when the system was ready to start an experimental session. The other

terminal was used to initialize experimental sessions and monitor the subject's performance.

Audio/video equipment. A JVC GX-5700U color video camera and a Sony Trinitron color monitor were used to monitor the subject during the experiment. No video recordings were made, and the subject was informed of the intended use of the camera. Also, a microphone placed in front of the subject was connected to the monitoring equipment to deliver audio signals from the subject's station to the experimenter's station.

A General Electric VHS video cassette recorder and a Sony Trinitron color monitor were used to provide the subject with a detailed explanation of the experiment and demonstration of two example tasks.

Experimental rooms. Three separate experimental rooms were used for this experiment. One experimental room was equipped with the computer system which was connected to the synthesizer and the visual display terminals. A second room was used to perform the experimental tasks and equipped with the synthesizer, audio/video equipment, and the telephone. The third room equipped with the monitoring system was used to monitor the subject. These rooms were physically insulated from each other and the subject's station was well protected from external noise sources. Figure 9 shows the layout of the set of equipment used in the experiment.

Information Databases

For search tasks, two different types of databases were developed (Beaudet, 1988): an 8x2 and a 2x6 databases. The hierarchical structures of these databases are presented in Figure 6 and Figure 7, respectively. In the Experiment 3, the 8x2 database structure was used since it was used in the first

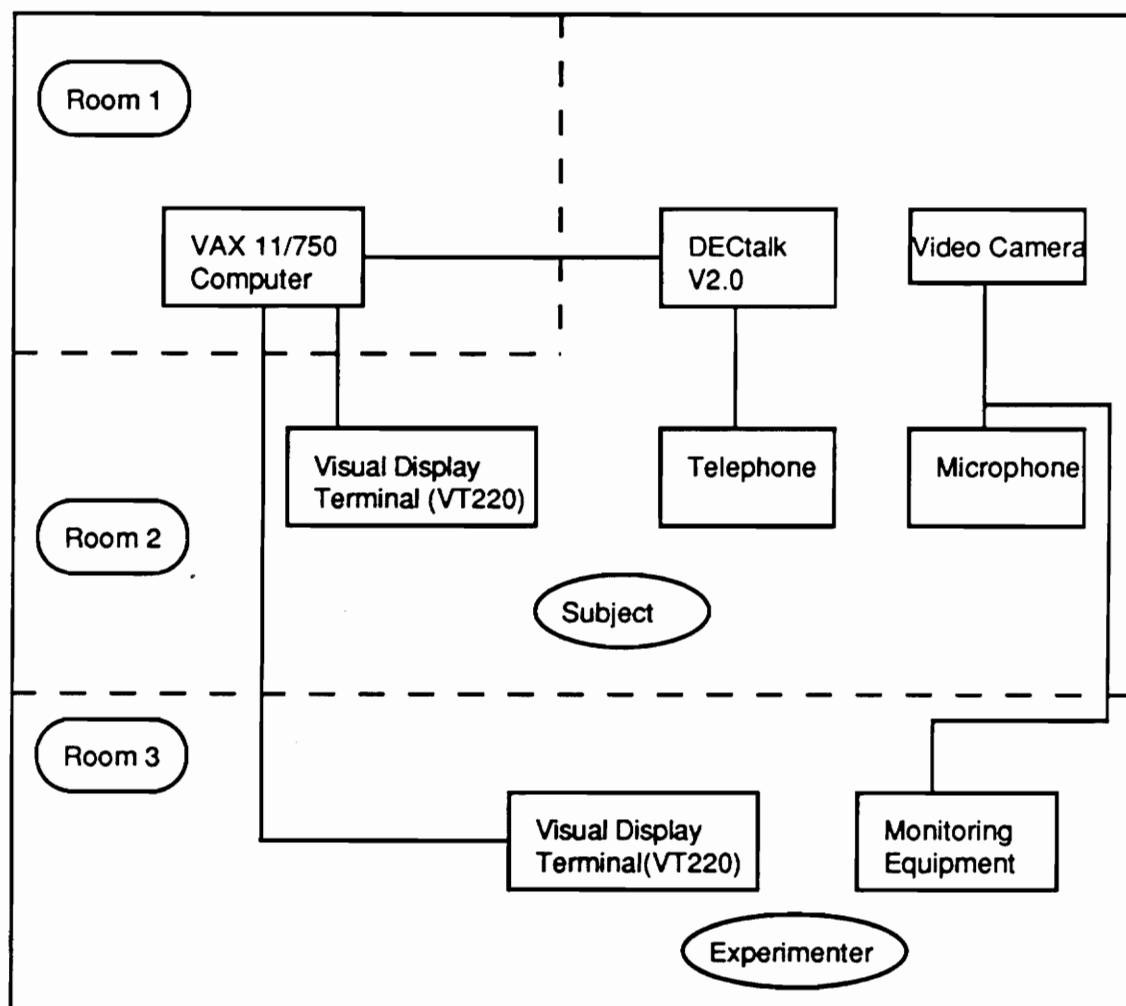


Figure 9. Configuration of experimental apparatus.

sequential experiment as a constant factor. In the Experiment 4, however, both types of databases were used depending upon the computer output for selecting additional design points.

Structures of databases. Both databases had the same number of alternatives at the bottom level of the hierarchy. That is, the 2x6 database had two alternatives within each of six menus. On the other hand, the 8x2 database had eight alternatives within each of two menus. Therefore, there were 64 nodes (2^6 or 8^2) at the bottom level of the hierarchy.

Each alternative within the hierarchical structures had a specific keyword (i.e., a group of related items) with different abstracts of product item categories. For example, "automotive" is a keyword for "tires" and "motor oils", etc. The top level of the hierarchy had general product categories (e.g., "automotive"). As the level of the hierarchy went downward from the top, the keywords for product items became more specific (e.g., "tires" for the second level of the hierarchy under "automotive" category).

Although both databases had the same number of alternatives at the bottom level, different types of keywords for the two databases were developed (Beaudet, 1988). In the 2x6 hierarchy, keywords were brief and usually a single word (e.g., dresses). In the 8x2 hierarchy, keywords were usually two- and sometimes three-word composite clauses (e.g., women's dresses). The 8x2 hierarchy had fewer menu items ($8^1 + 8^2 = 72$) than the 2x6 hierarchy ($2^1 + 2^2 + 2^3 + 2^4 + 2^5 + 2^6 = 126$). Therefore, menu items in the 8x2 hierarchy required more information than the 2x6 hierarchy. Based upon the information required to develop the databases, keywords were developed to minimize semantic ambiguity.

Product information messages. One unique product information message was developed for each target item, and hence there were 64 different information messages. Appendix H presents these product information messages. All information messages had the same sentence structure, i.e., "Modifier subject verb preposition modifier object". For example, "Spring mini-skirts are offered with matching pocketbooks" is an example of the information message that follows the stated rule. Each modifier, subject, and object were used only once in constructing the 64 information messages. All information messages were classified into the following four categories: Location, Price, Availability, or Information. Table 12 shows the structures of the information messages. Transcription accuracy was scored only for the first and the last two words of each sentence (Merva, 1988) since the middle parts of the sentences (i.e., verb and preposition) were repeated during the experiment.

Experimental Procedures

The outline of experimental procedures is presented in Table 13. At the beginning of the experimental session, each subject received a brief description of the experiment and was given an informed consent document to sign. Demographic information was collected for each subject. After the completion of the demographic questionnaire, subjects were asked if they had had experiences in hearing any synthetic speech. All subjects had not had any previous experience with synthetic speech similar to this system. As explained in the previous section, a hearing test was administered to determine if the subject had normal hearing.

Instruction session. Following the hearing test, an instruction session began. The instruction session consisted of four different subsessions: general

TABLE 12

Sentence Structures of Product Information Messages

Information Type	Format
Location:	Modifier subject is/are <u>in</u> modifier object. on near
Price:	Modifier subject is/are <u>by</u> modifier object. for Modifier subject are sold <u>by</u> modifier object. for
Availability:	Modifier subject is/are available <u>at</u> modifier object. by in with
Information:	Modifier subject are offered <u>on</u> modifier object. to for with Modifier subject is/are required <u>on</u> modifier object. to for within

TABLE 13
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	
Rest Period	
Targets 9-16	
Search Task	
Transcription Task	
Post Experimental Session	(~ 15 minutes)
Debriefing	
Payment and Dismissal	

orientation of this experiment, detailed experimental procedures, video instructions, and command function keypads (see Appendix G). All these instructions except the video instruction were given to the subject through the telephone speaker which spoke the instructions using the synthetic speech. The purpose of audio instructions was to familiarize the subject with the synthetic speech and also to reduce the possibility that the subject intentionally skip portions of the instruction. Written documents of instructions were also given to the subject for reference. During the instruction session, the subject was required to listen to the instruction carefully and was allowed to ask any questions to the experimenter. The following section describes a detailed procedures of the instruction session.

- (1) The subject was required to press the "On-Off" button on the telephone to activate the speaker, and then the telephone delivered a simple tone of the ready signal.
- (2) The subject made a phone call to the hypothetical department store by pressing the preassigned button.
- (3) The telephone information system started reading the general introduction of this experiment.
- (4) At the end of the introduction, the subject was required to press the space bar of the keyboard while the visual display terminal showing "press the space bar". Then, the second instruction session for detailed experimental procedures was spoken.
- (5) There were 17 steps of instructions for a single target search task or 20 steps of instructions for a multiple target search task. In a single target search task, the subject was required to find a single target,

while he/she was required to find two targets in a multiple target search task.

- (6) At the end of the second instruction subsession, the subject was allowed to ask questions on the experimental procedures.
- (7) The third subsession was a video instruction in which example tasks and experimental procedures were demonstrated.
- (8) After the video instruction, command function keypads for performing the search tasks were explained to the subject using synthetic speech.
- (9) At the end of the instruction session, the subject was required to explain the experimental tasks and procedures. The experimenter corrected any misinterpretations or misunderstandings.

Practice session. Two practice trials were administered to make sure that the subject understood the experimental tasks and procedures. The subject began the practice target searches by calling the information system explained in the previous section. After the practice targets were completed, any questions the subject might have were answered.

Experimental session. During the experimental session, the subject performed 16 trials of target searches and information transcriptions. Each trial consisted of a search task and a transcription task. After finishing a half of the 16 trials, the subject took a short rest (about five minutes) and began the second half of the trials. The detailed experimental procedures follow.

- (1) The visual display terminal displayed two steps of instructions for calling the information system. These were "pressing the On-Off button to activate the speaker", and "pressing the preassigned dial

- button". The subject called the information system according to these instructions.
- (2) The information system asked if the subject would like to have instructions. The subject could press the "#" key for the instructions or wait for a few seconds (approximately 2-5 seconds) to skip the instructions.
 - (3) The terminal then displayed the product(s) that was the target item(s). The instructions requested that the subject find the information message for a specific product item (or two items for multiple search tasks). For a single target search, the message would read "What is the information message for hand cream?". For a multiple target search, the target message would read "What are the information messages for telephones and cocktail dresses?". The message remained displayed during the entire target search.
 - (4) Fifteen seconds after the target was displayed, a "ready..." message was displayed below the target message to indicate that the search was about to begin.
 - (5) Two seconds after the ready message, a "Begin the search for the store item" message was displayed on the screen, and the information system spoke the first keyword in the main menu. The first store item in the main menu was "Furniture" for the 8x2 database.
 - (6) If the subject determined the first main menu (i.e., "Furniture") was the appropriate category for the target store item, then he/she was required to press "#" key on the telephone keypad to select that category. Then, the information system would speak the first store

item (i.e., "Recliner Chairs") under the "Furniture". On the other hand, if the subject determined the first item was not the appropriate category for the target product item, the subject would wait for three seconds or nine seconds to listen to the second item in the main menu. In this fashion, the subject selected the appropriate categories in descending order until the subject found the target item.

- (7) If the subject made an error, then two functions keys could be used to return to previous menus. The "*" key could be used to back up one menu level, while the "0" key could be used to return to the main menu. The subject was required to make a decision on which approach would take less time and keypresses to reach the target item.
- (8) If the subject located the target items by selecting the appropriate categories for the target item, the information system stated: "At store item keyword, press the "2" key to hear the information message." As soon as the subject pressed the "2" key on the telephone keypad, the information system spoke a sentence of information message about the target item. If the subject reached an incorrect item, the information system stated: "At store item keyword, continue searching". The subject would then have to leave the item by pressing either the "*" key or the "0" key on the telephone keypads and continue searching for the correct store item using the procedures described above.
- (9) When the correct item was found, the system spoke an information message followed by a prompt to the subject to "Begin transcription".

The subject transcribed the message using the computer keyboard. If he/she could not recognize or remember the entire information message, the best guess of the message was entered. If he/she made typing mistakes, arrow keys on the keyboard could be used to correct the mistakes. After finishing the transcription, the subject pressed the "return" key on the keyboard to record the transcription.

- (10) If the subject were performing a single target search task, the information system displayed the next target store item on the computer screen. The subject followed the same procedures described above until he/she finished a half of the target searches (eight target store items). After a short rest period (approximately five minute), the second half of the searches began.
- (11) In the case of a multiple search task, the computer screen displayed "ready..." and "Begin search for the second store item" immediately after the subject completed the transcription task. Two seconds later, the information system stated: "At store item, keyword, continue searching". The subject was then required to press either the "" key or the "0" key to return to the appropriate category. The subsequent procedures were the same as the previous steps described in the steps from (3) to (10).

Post Experimental session. After the experimental session, a subject debriefing was conducted, and questions by the subject were answered. A questionnaire (see Appendix I) was given to the subject to obtain any suggestions and feedbacks for improving the design of the telephone

information system. The subject was then thanked for his/her participation and paid upon the completion of the debriefing session.

Dependent Measures

Four dependent measures (i.e., TST, UAT, AKP, AND STA) were collected during the experiment. The description of these dependent measures is presented in Chapter 3.

Results

The metering software on the VAX computer automatically recorded the data collected during this experiment. But the average transcription accuracy scores were manually calculated. Using a data reduction package, the data resulting during the search tasks were averaged to obtain average performance measures since each subject performed 16 trials. That is, the data (four measures) resulting from the experiment were reduced to obtain an average score for each subject. Analysis of variance techniques were used to analyze the four dependent measures. The analyses of variance were performed with the IBM 3090 mainframe computer at Virginia Polytechnic Institute and State University using the Statistical Analysis System (SAS, 1986).

Total Search Time (TST)

Total search time varied from a minimum value of 29.6 seconds per trial to a maximum of 93.8 seconds per trial. The average total search time was 56.6 seconds per trial. The ANOVA included five main effects and ten two-way interactions. Table 14 shows the results of the ANOVA. As shown in Table 14, two main effects (SR and IT) were found to be statistically significant ($p < 0.05$).

On the other hand, there was no significant interaction effects. Since each significant main effects had only two levels, no post-hoc analyses were performed. Figure 10 shows the average values of these two significant main effects. When the speech rate was slow (i.e., 138 words per minute; 58.2 seconds of total search time) and the input timeout was long (i.e., 9 seconds; 77.8 seconds of total search time), the total search time was significantly longer than that of the other conditions (55.0 and 35.4 seconds of total search time, respectively) ($p < 0.05$). These two main effects will be used in building an integrated empirical model for total search time (see Chapter 5).

User Added Time (UAT)

User added time was calculated by subtracting minimum search time from total search time (see Chapter 3 for a detailed description). User added time was varied from a minimum of -1.3 seconds per trial to a maximum of 23.4 seconds per trial. The negative value means that the subject completed the search task faster than an average ideal user. The average user added time was 5.3 seconds per trial. The results of the analysis of variance showed that there was no significant effect at all as shown in Table 15. That is, the five main effects and the ten two-way interactions did not affect the user added time. This result means that the integrated empirical model for the user added time will not include any independent variables resulting from this experiment.

User Added Keypresses (AKP)

User added keypresses were calculated by subtracting the minimum number of keypresses from the total number of keypresses used to find a target (see Chapter 3 for a detailed description). The minimum number of keypresses for the 8x2 database structure was three for a single target search task and

TABLE 14

ANOVA Summary Table for Total Search Time (TST)

Source	df	SS	MS	F	p
<u>Main Effects</u>					
SR(speech rate)	1	116.9657	116.9657	6.84	0.0135*
IT(input timeout)	1	21619.3183	21619.3183	1263.77	0.0001*
NT(no. of targets)	1	2.6428	2.6428	0.15	0.6969
TV(type of voice)	1	23.8107	23.8107	1.39	0.2468
SEX(sex of subject)	1	1.3085	1.3085	0.08	0.7839
<u>Interactions</u>					
SRxIT	1	1.0648	1.0648	0.06	0.8046
SRxNT	1	34.5968	34.5968	2.02	0.1647
SRxTV	1	0.0778	0.0778	0.00	0.9466
SRxSEX	1	52.3942	52.3942	3.06	0.0897
ITxNT	1	24.5917	24.5917	1.44	0.2393
ITxTV	1	15.5326	15.5326	0.91	0.3473
ITxSEX	1	15.7197	15.7197	0.92	0.3449
NTxTV	1	17.8596	17.8596	1.04	0.3146
NTxSEX	1	24.7466	24.7466	1.45	0.2379
TVxSEX	1	4.0142	4.0142	0.23	0.6314
<u>Error</u>	32	547.4233	17.2433		
<u>Total</u>	47	22502.0672			

* Significant at $\alpha = 0.05$

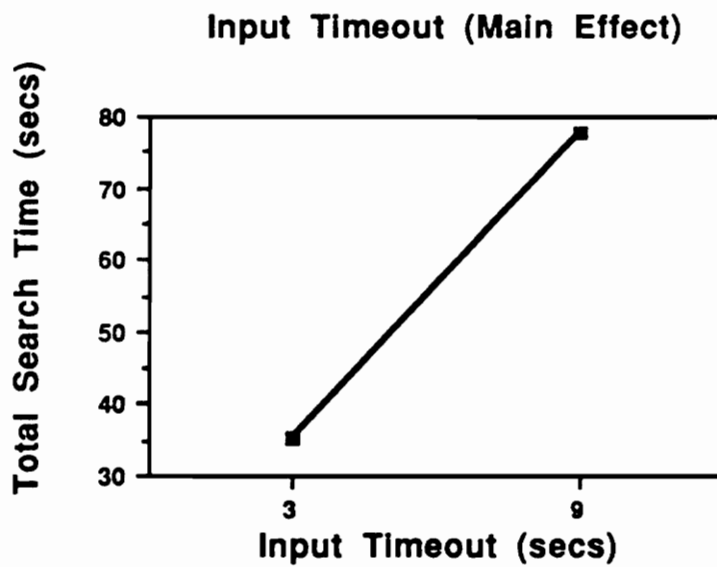
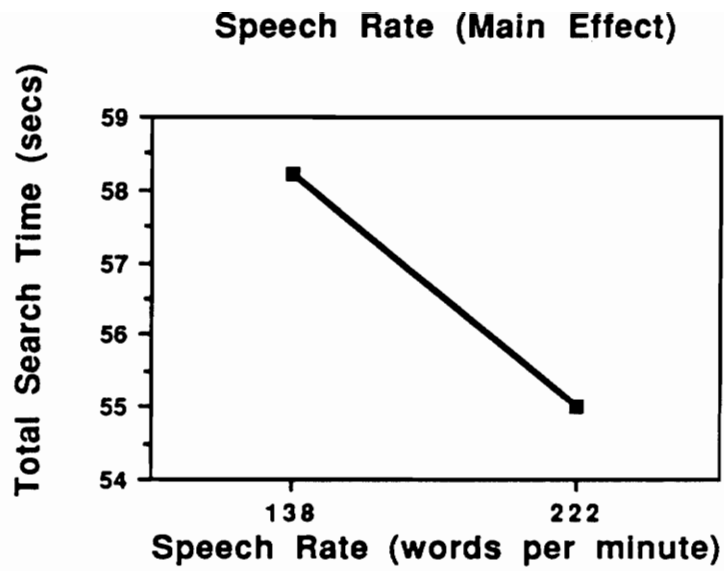


Figure 10. Significant main effects for total search time.

TABLE 15

ANOVA Summary Table for User Added Time (UAT)

Source	df	SS	MS	F	p
<u>Main Effects</u>					
SR(speech rate)	1	32.2703	32.2703	1.89	0.1792
IT(input timeout)	1	49.8882	49.8882	2.92	0.0974
NT(no. of targets)	1	0.2985	0.2985	0.02	0.8957
TV(type of voice)	1	23.6813	23.6813	1.38	0.2480
SEX(sex of subject)	1	1.2822	1.2822	0.01	0.7860
<u>Interactions</u>					
SRxIT	1	1.0863	1.0863	0.06	0.8027
SRxNT	1	17.0468	17.0468	1.00	0.3257
SRxTV	1	0.0873	0.0873	0.01	0.9435
SRxSEX	1	52.4653	52.4653	3.07	0.0895
ITxNT	1	0.3511	0.3511	0.02	0.8870
ITxTV	1	15.5667	15.5667	0.91	0.3473
ITxSEX	1	15.6214	15.6214	0.91	0.3463
NTxTV	1	18.0161	18.0161	1.05	0.3125
NTxSEX	1	25.0346	25.0346	1.46	0.2352
TVxSEX	1	4.0933	4.0933	0.24	0.6281
<u>Error</u>	32	547.4256	17.1070		
<u>Total</u>	47	804.2148			

* Significant at $\alpha = 0.05$

seven for a multiple target search. User added keypresses varied from a minimum of zero per trial to a maximum of 0.69 per trial. The average user added keypress was 0.30 per trial. The analysis showed that only one main effect (NT) was found to be significant ($p < 0.05$) and there were no significant two-way interaction effects. Table 16 shows the results of the ANOVA and Figure 11 presents the mean values of the significant main effect. The user added keypress with a multiple target search was significantly greater than that with a single target search.

Transcription Average (STA)

In the second part of the experiment, the subject was required to transcribe an information message spoken by the speech synthesizer. Since only the first two and the last two words of each sentence were scored, the highest number of transcription scores was four words per trial and the lowest number was zero. Transcription average scores varied from a minimum of 3.13 to a maximum of 3.96. The average score was 3.53. Table 17 shows the results of the ANOVA. As shown in Table 17, only one main effect (SR) and one two-way interaction (IT x SEX) were significant ($p < 0.05$). Figure 12 and Figure 13 illustrate the significant main effect and the interaction effect, respectively. The transcription accuracy with the lower speech rate (i.e., 138 words per minute) was higher than that with the higher rate (222 words per minute).

Since the interaction effect was significant, Newman-Keuls post-hoc analyses were performed. When the input timeout was set at 3 seconds, male subjects transcribed the information messages more accurately than female subjects ($p < 0.05$). However, there was no significant difference in

transcription accuracy between male and female subjects with the longer input timeout (i.e., 9 seconds). While male subjects did not show any difference in transcription accuracy between the shorter and the longer input timeout ($p > 0.05$), female subjects showed more accurate transcription with the longer input timeout than that with the shorter input timeout ($p < 0.05$). Table 18 shows the results of the post-hoc analyses.

TABLE 16

ANOVA Summary Table for User Added Keypresses (AKP)

Source	df	SS	MS	F	p
<u>Main Effects</u>					
SR(speech rate)	1	0.0450	0.0450	1.42	0.2427
IT(input timeout)	1	0.0251	0.0251	0.79	0.3808
NT(no. of targets)	1	0.9933	0.9933	31.30	0.0001*
TV(type of voice)	1	0.0222	0.0222	0.70	0.4088
SEX(sex of subject)	1	0.0309	0.0309	0.97	0.3315
<u>Interactions</u>					
SRxIT	1	0.0149	0.0149	0.47	0.4985
SRxNT	1	0.0171	0.0171	0.54	0.4688
SRxTV	1	0.0759	0.0759	2.39	0.1318
SRxSEX	1	0.5706	0.5706	1.80	0.1894
ITxNT	1	0.0448	0.0448	1.41	0.2434
ITxTV	1	0.0108	0.0108	0.34	0.5643
ITxSEX	1	0.0811	0.0811	2.56	0.1198
NTxTV	1	0.0090	0.0090	0.29	0.5971
NTxSEX	1	0.0759	0.0759	2.39	0.1318
TVxSEX	1	0.0010	0.0010	0.03	0.8603
<u>Error</u>	32	1.0156	0.0317		
<u>Total</u>	47	2.5197			

* Significant at $\alpha = 0.05$

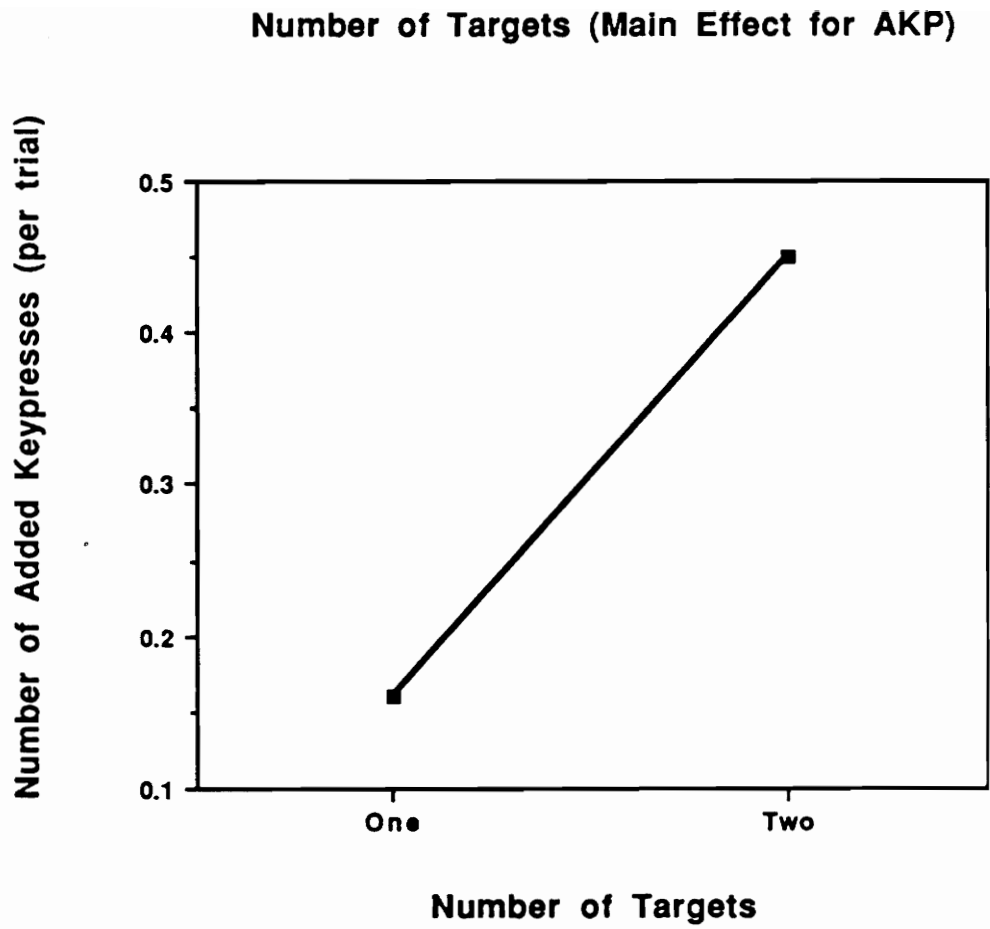


Figure 11. Significant main effect for user added keypresses.

TABLE 17

ANOVA Summary Table for Transcription Averages (STA)

Source	df	SS	MS	F	p
<u>Main Effects</u>					
SR(speech rate)	1	0.7191	0.7191	21.99	0.0001*
IT(input timeout)	1	0.0007	0.0007	0.02	0.8820
NT(no. of targets)	1	0.0183	0.0183	0.56	0.4597
TV(type of voice)	1	0.0013	0.0013	0.04	0.8431
SEX(sex of subject)	1	0.0294	0.0294	0.90	0.3503
<u>Interactions</u>					
SRxIT	1	0.0264	0.0264	0.81	0.3759
SRxNT	1	0.0208	0.0208	0.64	0.4306
SRxTV	1	0.0509	0.0509	1.56	0.2213
SRxSEX	1	0.0052	0.0052	0.16	0.6925
ITxNT	1	0.0098	0.0098	0.30	0.5870
ITxTV	1	0.0029	0.0029	0.09	0.7666
ITxSEX	1	0.2286	0.2286	6.99	0.0126*
NTxTV	1	0.0264	0.0264	0.81	0.3759
NTxSEX	1	0.1238	0.1238	3.79	0.0605
TVxSEX	1	0.0117	0.0117	0.36	0.5536
<u>Error</u>	32	1.0462	0.0327		
<u>Total</u>	47	2.3215			

* Significant at $\alpha = 0.05$

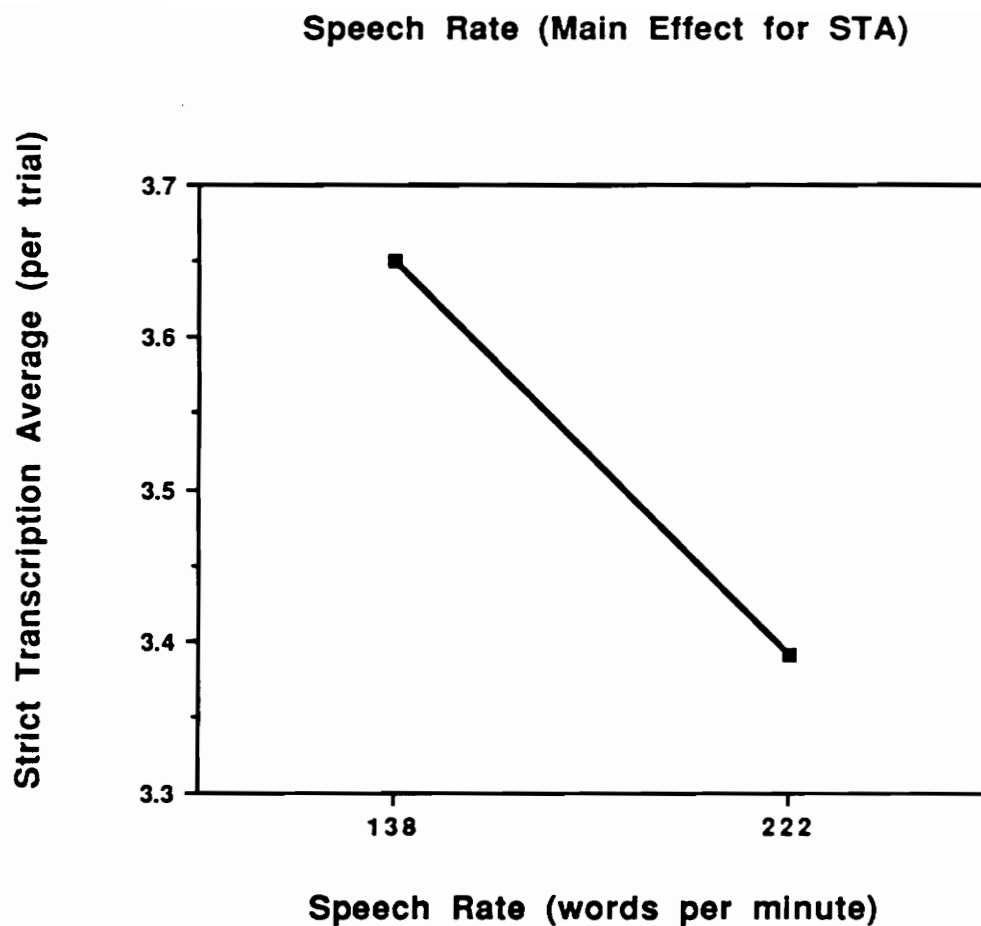


Figure 12. Significant main effect for transcription average.

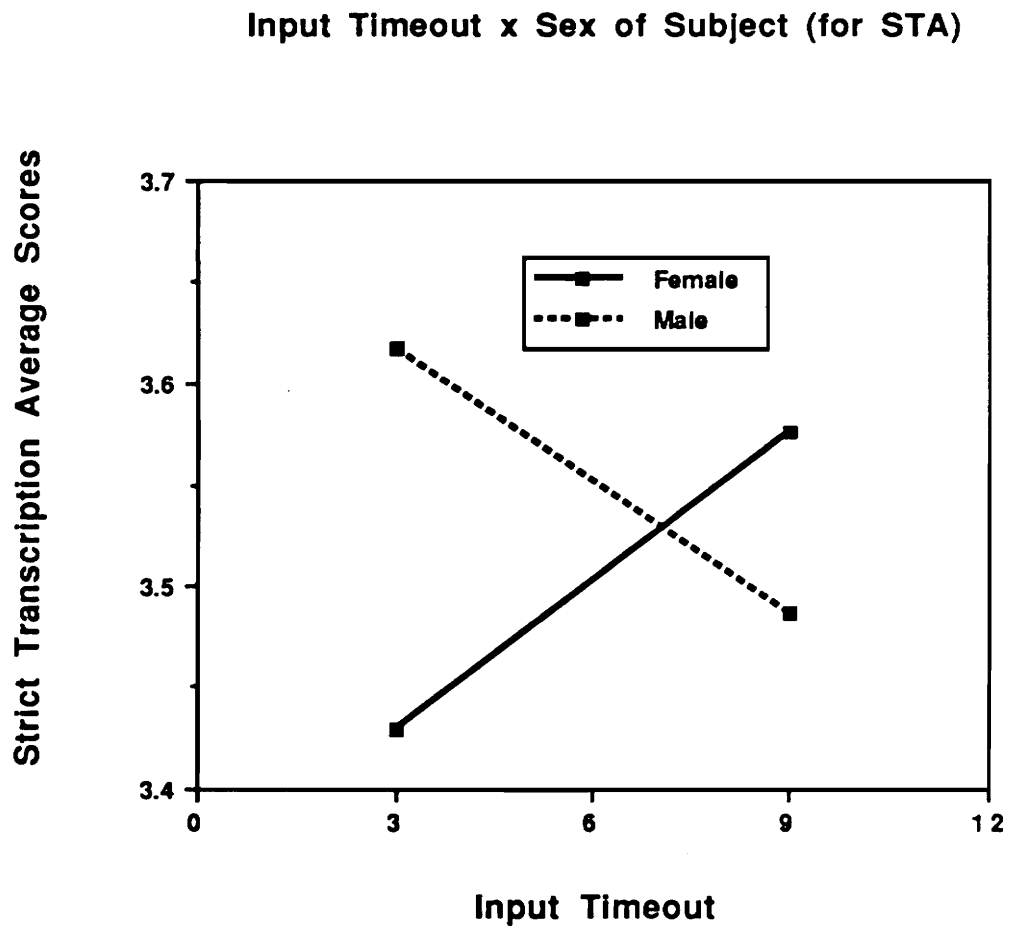


Figure 13. Significant interaction for transcription average.

TABLE 18
Results of Newman-Keuls Tests for ITxSEX Interaction

Constant	Tested Variable	Mean	
IT=3*	SEX=Female	3.42	A**
	SEX=Male	3.61	B
IT=9	SEX=Female	3.57	A
	SEX=Male	3.46	A
SEX=Female	IT=3	3.42	A
	IT=9	3.57	B
SEX=Male	IT=3	3.61	A
	IT=9	3.46	A

* seconds

** Means with the same letter are not significantly different from one another at $p < 0.05$

CHAPTER 5

DATA BRIDGING

Background

Following the sequential studies, a validation study of the results needs to be conducted using the combined results from the earlier studies. Since each individual study in the sequential research strategy investigated only a subset of the independent variables while the remaining variables were held constant, the data set resulting from each study describes the functional relationships only among the independent variables investigated in each individual study. That is, the results of each study form an empirical data set which provides limited design implications because usually only a smaller number of variables were investigated in each study. These individual data sets need to be combined together for describing the functional relationships provided by all factors considered in each study.

In combining the data sets resulting from the several sequential studies, two fundamental assumptions should be made (see Chapter 1). First, only two-way interactions which have not been investigated in the previous studies are examined in the integration procedures. Second, a second-order polynomial expression is used to approximate the complex relationships among the independent variables. Based upon these two assumptions, Han, Williges, and

Williges (1990) described procedures and methods necessary to combine several data sets. Appendix J presents a theoretical background of their procedures and methods.

Uninvestigated Interactions

Background

To investigate a two-factor interaction, at least four different design points are required as long as each level of a factor is related to at least two different levels of the other factor. That is, it is impossible to estimate a two-factor interaction effect with less than four different design points and even with four design points if more than two different design points (i.e., three or four) are on the same line within the experimental space created by the two factors. Figure 14 illustrates some examples where it is impossible to estimate the two-factor interaction effect between X_1 and X_2 .

To find additional design points required to build a second-order empirical model, the experimenter must determine the experimental region of interest. When each experimental variable is defined by an upper and lower bound, the experimental region is a hypercube (i.e., a cuboidal region). If experimentation in the corners of the hypercube is impossible or unnecessary, it may be appropriate to employ a spherical design in which a hypersphere, cutting off the corners of the hypercube, is used. For a spherical region of interest, additional design points should be determined in such a way that they are located inside a hypersphere generated by the experimental variables. For example, if the experimenter used a CCD for a spherical region of interest (i.e.,

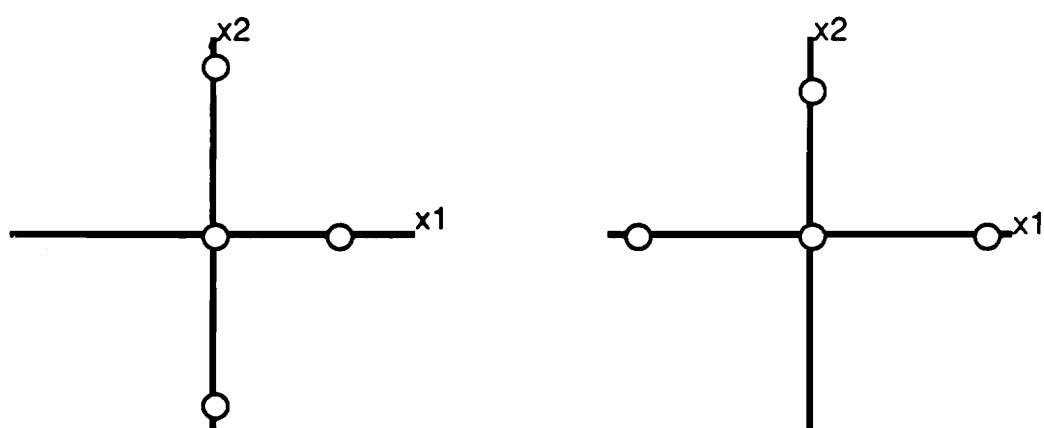


Figure 14. Some examples where it is impossible to investigate a two-factor interaction with four design points.

$\alpha > 1.0$), then an additional design point should be selected within the region generated by the factorial portion in the CCD since the information required to estimate interaction effects is derived from the factorial portion.

Identifying Uninvestigated Interactions

Given the two assumptions, the first step to identify newly created interactions (i.e., interactions uninvestigated in each study) is to list the main effects and their two-factor interactions present in each study. In Experiment 1 (Wu, 1989), four main effects (SR, IT, MUS, and AGE) and six two-factor interaction (SR x IT, SR x MUS, SR x AGE, IT x MUS, IT x AGE, and MUS x AGE) were investigated along with the pure quadratic effects of the four independent variables. Therefore, the other interactions with the four variables could not be estimated using the results of the first sequential experiment since all other factors were held constant during the experiment. That is, the other interactions were confounded with the manipulated factors so that it was impossible to estimate their effects without additional design points. Five main effects (SR, IT, DB, FB, and WG) and their two-factor interactions (SR x IT, SR x DB,...,FB x WG) were investigated in Experiment 2 (Martin, et al., 1990). Finally, five main effects (SR, IT, NT, TV, and SEX) and their two-factor interactions (SR x IT, SR x NT, ..., TV x SEX) were investigated in Experiment 3. Table 19 and Table 20 show respectively the main effects and two-factor interactions investigated during the three sequential experiments.

As shown in Table 19, all main effects of the ten independent variables resulting from the screening study were orthogonally estimated during the three sequential studies after the completion of Experiment 3. However, all possible combinations of two-factor interactions among the ten factors were not

TABLE 19
Main Effects Investigated in Each Sequential Study

Experiment	Main Effect Investigated
Experiment 1 (Wu, 1989)	SR*, IT, MUS, AGE
Experiment 2 (Martin, 1990)	SR, IT, WG, DB, FB
Experiment 3	SR, IT, NT, TV, SEX
* SR : Speech Rate IT : Input Timeout TV : Type of Voice MUS : Background Music Level AGE : Age of Subject WG : Wallet Guide DB : Database Organization FB : Feedback NT : Number of Target Items SEX : Sex of Subject	

TABLE 20

Two-factor Interactions Investigated in Each Sequential Study

Experiment	Two-factor Interaction Investigated	Number of Interactions
Experiment 1 (Wu, 1989)	SR* x IT, SR x MUS, SR x AGE, IT x MUS, IT x AGE, MUS x AGE	6
Experiment 2 (Martin, 1990)	SR x IT, SR x WG, SR x DB, SR x FB, IT x WG, IT x DB, IT x FB, WG x DB, WG x FB, DB x FB	9
Experiment 3	SR x IT, SR x NT, SR x TV, SR x SEX, IT x NT, IT x TV, IT x SEX, NT x TV, NT x SEX, TV x SEX	9
		Total 24
* SR : Speech Rate IT : Input Timeout MUS : Background Music Level AGE : Age of Subject TV : Type of Voice WG : Wallet Guide DB : Database Organization FB : Feedback NT : Number of Target Items SEX : Sex of Subject		

investigated. That is, the number of two-factor interactions among the ten variables is 45 ($_{10}C_2 = 45$), whereas the number of two-factor interactions investigated during the three studies is 24 as shown in Table 20. Therefore, the remaining 21 two-factor interactions should be investigated to build a complete second-order empirical model. Table 21 shows the list of the 21 uninvestigated interactions which are newly created by combining the individual data sets into one integrated data set for building a second-order empirical model. These uninvestigated interactions cannot be estimated even though all three data sets resulting from each sequential study are combined together into one integrated data set since they are completely confounded with other factors. Additional design points should be added to the combined data set to estimate the effects of the uninvestigated interactions.

Additional Design Points

Background

For each uninvestigated interaction, it is necessary to provide one additional design point to provide four different combinations of design points required to estimate the interaction since the combined data set provides three different design points for each uninvestigated interaction. In other words, 21 additional design points should be added to the combined data set resulting from the three studies to estimate the 21 uninvestigated interactions. For example, the uninvestigated interaction between X_3 and X_5 (i.e., X_3X_5) has three different design points in the combined data set. The three design points are $(X_3, X_5) = (1, 1)$, $(-1, 1)$, and $(-1.4142, -1)$. Therefore, only one additional

TABLE 21

Uninvestigated Two-factor Interactions During the Sequential Studies

MUS x DB*, MUS x FB, MUS x WG, MUS x NT, MUS x TV, MUS x SEX, AGE x DB, AGE x FB, AGE x WG, AGE x NT, AGE x TV, AGE x SEX, DB x NT, DB x TV, DB x SEX, FB x NT, FB x TV, FB x SEX, WG x NT, WG x TV, WG x SEX	
Total: 21 Interactions	
* SR : Speech Rate	WG : Wallet Guide
IT : Input Timeout	DB : Database Organization
TV : Type of Voice	FB : Feedback
MUS : Background Music Level	NT : Number of Target Items
AGE : Age of Subject	SEX : Sex of Subject

design point needs to be added to provide four different design points required to estimate the interaction between X_3 and X_5 .

In selecting an additional design point, it is necessary to develop decision criteria to obtain desirable statistical properties such as prediction variances and variances of coefficients since there is a great deal of possible candidates for the fourth design point. For example, in selecting the fourth design point for the investigation of X_3X_5 , an arbitrary number between -1 and +1 can be assigned to the level of X_3 since X_3 is a continuous variable. It is generally known, however, that the variance of coefficient of a model item in a linear regression equation is minimum when the factor levels are set at its extremes within its region of interest (Dykstra, 1971). Therefore, $(X_3, X_5) = (+1, -1)$ provides the optimal additional design point among the infinite number of candidates since the information for estimating an interaction effect is stemming from the factorial portion rather than the axial or center points in a spherical CCD.

Likewise, the fourth design points of all uninvestigated interactions can be determined based on the criterion described above. Table 22 shows the fourth design points for all uninvestigated interactions in the telephone information study.

As shown in Table 22, the additional design point for each uninvestigated interaction can be expressed as the levels between the two factors. However, Table 22 does not show which specific treatment combination of all ten factors should be selected as the final design point. For example, the fourth design point for X_3X_5 can be any one of 2^8 combinations generated by the other factors (i.e., $X_1, X_2, X_4, X_6, X_7, X_8, X_9$, and X_{10}).

TABLE 22

The Fourth Design Point for the Uninvestigated Interactions in the Telephone Information Study

Uninvestigated Interaction	X ₁	X ₂	X ₃	X ₄	Factor X ₅	Level X ₆	X ₇	X ₈	X ₉	X ₁₀
X ₃ X ₅ *			+1		-1					
X ₃ X ₆			+1			+1				
X ₃ X ₇			+1				+1			
X ₃ X ₈			+1					-1		
X ₃ X ₉			+1						+1	
X ₄ X ₅				+1	-1					
X ₄ X ₆				+1		+1				
X ₄ X ₇				+1			+1			
X ₄ X ₈				+1				-1		
X ₄ X ₉				+1					+1	
X ₅ X ₈					-1			-1		
X ₅ X ₉					-1				+1	
X ₆ X ₈						+1		-1		
X ₆ X ₉						+1			+1	
X ₇ X ₈							+1	-1		
X ₇ X ₉							+1		+1	

* X₃=MUS (+1=61.1 dBA), X₄=AGE (+1=Middle Age), X₅=DB (-1=2x6),
X₆=FB (+1=Available), X₇=WG (+1=Available), X₈=TV (-1=Betty), X₉=NT (+1=Two)

Decision criteria for selecting the fourth design point should be determined in order to choose the best design point among the 2^8 candidates.

Statistical criteria for selecting additional design points. When the experimenter wants to build an empirical model for describing the functional relationships between independent variables and their responses, the following linear model approximation is usually used (Myers, 1990);

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{Eq. (2)}$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations, $\mathbf{X}$ is an $n \times k$ of constants, $\boldsymbol{\beta}$ is the $k \times 1$ vector of coefficients to be estimated, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of independently and identically distributed random variables, each having mean 0 and unknown variance σ^2 . There are a variety of statistical criteria for constructing optimal designs to fit the linear model. One of the most popular is the maximization of $|\mathbf{X}'\mathbf{X}|$, i.e., minimization of the generalized variance of the estimates of the coefficients (Mitchell, 1974a). One of the attractive features of the $|\mathbf{X}'\mathbf{X}|$ criterion is that designs which are optimal with respect to it are invariably good in many respects (e.g., low variances for the coefficients, low correlations among the coefficients, and low maximum prediction variance over the experimental region of interest) (Mitchell, 1974b). In other words, the maximization of $|\mathbf{X}'\mathbf{X}|$ means the minimization of the volume of confidence intervals generated by the model.

In augmenting additional design points with existing data, the $|\mathbf{X}'\mathbf{X}|$ criterion has been used to find appropriate design points (Dykstra, 1971; Evans, 1979; Hebble and Mitchell, 1972; Mitchell, 1974a, 1974b). This criterion is based on a theory that describes the relationship between the maximization of $|\mathbf{X}'\mathbf{X}|$ and prediction variance within the design region of interest. That is, the

additional design points should be selected at the point(s) where the prediction variance is maximum. Dykstra (1971) proved that the design point can be obtained by maximizing the $[\mathbf{X}'\mathbf{X}]$. However, this procedure requires the experimenter to specify an initial set of candidates for the additional design points. In choosing specific combinations of the candidates, the experimenter should be guided by the model which he wants to build (Dykstra, 1971). For a first-order model candidates should be selected at the extremes of the experimental space, so that only corner points are specified as candidates. For a second-order model the list of candidates should include axial points and a center point, in addition to corner points.

Based on these statistical criteria, additional design points were selected to investigate interactions uninvestigated during the three sequential experiments (i.e., Experiment 1, 2, and 3). The following sections describe these procedures in detail.

Selecting Additional Design Points

Preliminary analyses. Before selecting additional design points necessary to investigate the 21 interactions, several analyses were conducted to determine the candidate design points. First, it was necessary to list significant effects resulting from the sequential studies since the integrated empirical models would include only significant effects as model items. That is, since all the main effects were examined orthogonally, it was not necessary to include insignificant effects in the integrated models. Likewise, there was no need to include insignificant two-factor interactions found in each sequential study. Detailed descriptions of the results and analyses of the sequential studies are presented in Chapter 3 and Chapter 4. Table 23 summarizes

significant effects (i.e., significant main effects and two-factor interactions) for the four dependent measures (i.e., TST, UAT, AKP, and STA). As shown in Table 23, six main effects (SR, IT, MUS, AGE, DB, and WG) and four two-way interactions (ITxMUS, ITxDB, ITxWG, and DBxWG) will serve as the existing model items in building an empirical model for the TST. Likewise, six main effects (SR, IT, MUS, AGE, DB, and WG) and three two-way interactions (ITxDB, ITxWG, and DBxWG) will be used to develop the functional relationship for UAT, and so forth.

Second, the level of AGE factor was reexamined. In the screening study (Beaudet and Williges, 1988), Experiment 2 (Martin, et al., 1990) and Experiment 3, the AGE factor was treated as a two-level discrete factor (i.e., college age and middle age). On the other hand, Experiment 1 (Wu, 1989) manipulated the AGE factor as a continuous factor. In combining the three data sets resulting from the three studies, it was necessary to treat the AGE factor as a two-level discrete factor. The AGE factor was recoded as a two-level discrete factor in combining the data sets.

Third, an analysis of the levels of factors manipulated in the three experiments was performed. As shown in Table 6 and Table 7, the SEX variable was not controlled in Experiment 1 (Wu, 1989) and Experiment 2 (Martin, et al., 1990). It was possible to examine the interaction effects between the SEX and the other variables which had not been investigated since the combined data set provided information necessary to examine these interaction effects. In other words, since all the main effects and interaction effects were not significant in the previous studies (see Table 19 and Table 20), the combined data set provided sufficient degrees of freedom necessary to

TABLE 23

Significant Effects Resulting From the Sequential Studies

Dependent Measure	Main Effect	Interaction Effect
TST*	SR**, IT, MUS, AGE, DB, WG	ITxMUS, ITxDB, ITxWG, DBxWG
UAT	SR, IT, MUS, AGE, DB, WG	ITxDB, ITxWG, DBxWG
AKP	MUS, AGE, DB, WG	ITxWG, DBxWG
STA	SR, MUS, AGE	ITxSEX, MUSxMUS

* TST: Total Search Time

AKP: User Added Keypresses

UAT: User Added Time

STA: Transcription Average

** SR : Speech Rate

IT : Input Timeout

TV : Type of Voice

MUS : Background Music Level

AGE : Age of Subject

WG : Wallet Guide

DB : Database Organization

FB : Feedback

NT : Number of Target Items

SEX : Sex of Subject

examine these additional interaction effects. These interactions included: MUSxSEX, AGExSEX, DBxSEX, FBxSEX, and WGxSEX.

Separate regression analyses for each dependent measure were used to analyze these interactions. Several statistical criteria were used as a means of determining the validity of examining these interactions. These included the variance inflation factor (VIF) and the ratio of maximum eigen value to minimum eigen value of the **X** matrix (Myers, 1990) since there could be serious multicollinearity among the model items. The VIF value should be less than 10 and the ratio should be less than 1000 in order to determine an insignificant model item (Myers, 1990) when there is serious multicollinearity. The results of the analyses showed that all tests satisfied these statistical criteria. Partial F-tests were used to determine if these interactions were significant ($\alpha=0.05$). All significant model items resulting from the previous studies were used in conducting the partial F-tests. For example, to test the interaction effect between MUS and SEX for TST, the six main effects and the four interaction effects shown in Table 23 served as the existing model items. If the p-value of the interaction between MUS and SEX is greater than 0.05, then it means that this interaction is not significant under the presence of the six main effects and the four interaction effects. The results of the analyses showed that all five interactions (i.e., MUS x SEX, AGE X SEX, DB x SEX, FB x SEX, and WG x SEX) for all dependent measures were found to be insignificant ($p > 0.05$). Table 24 showed the p-values of these interaction effects. These interactions were not used in determining additional design points and hence, were not used in developing empirical models. Therefore, the number of uninvestigated

TABLE 24

Analysis Results of SEX Interaction Effects

Dependent Measure	<i>p</i> value of Interaction Between SEX and				
	MUS**	AGE	DB	FB	WG
TST*	0.2371	0.5571	0.5523	0.2370	0.6206
UAT	0.2996	0.9696	0.6627	0.2619	0.6963
AKP	0.7029	0.3586	0.7723	0.4565	0.5350
STA	0.4484	0.3206	0.3113	0.9768	0.9284

* TST: Total Search Time

AKP: User Added Keypresses

UAT: User Added Time

STA: Transcription Average

** MUS : Background Music Level

DB : Database Organization

WG : Wallet Guide

AGE : Age of Subject

FB : Feedback

SEX : Sex of Subject

interactions was reduced from 21 to 16. Table 25 lists the reduced set of uninvestigated interactions.

Development of a computer program. In selecting additional design points for the uninvestigated interactions specified in Table 25, the $|X'X|$ criterion was used. In constructing the X matrix, a trade-off was made between the choice of the best design points and data collection economy. The result was that all combinations of significant effects were used in constructing the X matrix across the dependent measures. That is, seven significant main effects, seven interactions, and one pure quadratic effect were used to construct the X matrix for the four dependent measures. It might not be the best solution in selecting the appropriate design point for each dependent measure since all dependent measure had different model items (see Table 23). Detailed discussions on the trade-off analysis are presented in Chapter 8.

Since all possible candidates for estimating an uninvestigated interaction can be determined using the theory described in the background section, the value of $|X'X|$ of each candidate was calculated and compared with each other to select a point providing the maximum value of $|X'X|$. A computer program (see Appendix K) was developed to determine the best design point for each uninvestigated interaction in a systematic manner since there were 2^8 candidate design points for each uninvestigated interaction. The algorithm of the computer program was developed by combining the procedures proposed by Dykstra (1971) and Mitchell (1974a, 1974b).

Step 1. Select an uninvestigated interaction among the set of 16 uninvestigated interactions.

TABLE 25
Reduced Set of Uninvestigated Two-factor Interactions

MUS x DB*, MUS x FB, MUS x WG, MUS x NT, MUS x TV,
AGE x DB, AGE x FB, AGE x WG, AGE x NT, AGE x TV,
DB x NT, DB x TV, FB x NT, FB x TV,
WG x NT, WG x TV

Total: 16 Interactions

* SR : Speech Rate	WG : Wallet Guide
IT : Input Timeout	DB : Database Organization
TV : Type of Voice	FB : Feedback
MUS : Background Music Level	NT : Number of Target Items
AGE : Age of Subject	SEX : Sex of Subject

- Step 2. Set the levels of each factor in the interaction at the missing corner to provide four different design points (see Table 22).
- Step 3. List all possible combinations of candidates for the uninvestigated interaction (2^8 treatment combinations).
- Step 4. Eliminate candidates which are overlapped with the existing design points.
- Step 5. Calculate the value of $|X'X|$ for each candidate.
- Step 6. Select the next interaction to be investigated.
- Step 7. Repeat Step 2 to Step 7 until the design points for all uninvestigated interactions are selected.
- Step 8. Select the candidate(s) which provides the maximum value of $|X'X|$.

Based on this algorithm, the computer program consisted of five sections: reading input data files (existing model items, candidate design points, and uninvestigated interactions); eliminating existing design points from the set of candidate design points; adding one candidate design point to the existing model items; calculating an $X'X$ matrix; and calculating the determinant value of the $X'X$ matrix. The existing model items contained all significant effects resulting from Experiment 1, 2, and 3. For the candidate design points, all possible factorial combinations generated by the ten independent variables were used. There were 1024 candidate design points since the level of each factor was set either at +1 or -1 coded value (i.e., $2^{10} = 1024$). Existing design points in the set of candidate design points were then eliminated since the set of candidate design points also included the existing design points which did not provide the fourth design point necessary to

investigate each uninvestigated interaction. Each candidate design point was added to the existing model items to calculate an $X'X$ matrix. Determinant values of $X'X$ matrices were then calculated using IMSL subroutines (called LFTRG and LFDRG). This procedure was repeated until all determinant values of the candidate design points for each uninvestigated interaction were obtained. The FORTRAN language was used to code the computer program. Appendix K presents the computer program list.

Selection procedures. The next step to select the first additional design point was to compare the maximum determinant values of the $X'X$ matrices and select the maximum of the maximum values as the final design point since there was a maximum value(s) for each uninvestigated interactions. The maximum of the maximum criterion was selected because it seemed to improve the statistical properties (i.e., maximum prediction variance and variances of coefficients) in an optimal manner. Table 26 shows the maximum determinant values of the candidates for the 16 uninvestigated interactions. As shown in Table 26, the candidate design point for the interaction between MUS and TV had the maximum value among the candidates. The specific factor level of the design point selected was (SR, IT, MUS, AGE, DB, FB, WG, TV, NT, SEX) = (138 wpm, 9 secs, 61.1 dBA, College Age, 2x6, Available, Available, Betty, Two, Male). This design point was selected to estimate the interaction effects.

However, this design point did not provide the factor level combination necessary to investigate the uninvestigated interactions such as AGExDB, AGExFB, AGExWG, AGExNT, and AGExTV since a college age subject was selected. In other words, the factor level combination necessary to investigate each uninvestigated interaction included (MUS, AGE, DB, FB, WG, TV, NT) =

TABLE 26

Maximum $|X'X|$ Value for Each Uninvestigated Interaction

(First Design Point)

Uninvestigated Interaction	Max $ X'X $ Value
$X_3X_5^*$	0.9681×10^{29}
X_3X_6	0.6733×10^{30}
X_3X_7	0.9681×10^{29}
X_3X_8	$0.6929 \times 10^{30**}$
X_3X_9	0.9681×10^{29}
X_4X_5	0.6644×10^{29}
X_4X_6	0.3479×10^{30}
X_4X_7	0.6644×10^{29}
X_4X_8	0.3532×10^{30}
X_4X_9	0.6644×10^{29}
X_5X_8	0.3536×10^{30}
X_5X_9	0.6644×10^{29}
X_6X_8	0.5877×10^{30}
X_6X_9	0.3479×10^{30}
X_7X_8	0.3536×10^{30}
X_7X_9	0.6644×10^{29}
* X_3=MUS, X_4=AGE, X_5=DB, X_6=FB, X_7=WG, X_8=TV, X_9=NT ** Maximum Value	

$(+1, +1, -1, +1, +1, -1, +1) = (61.1 \text{ dBA, Middle AGE, 2x6, Available, Available, Betty, Two})$ since the other factor level combinations were included in the combined data set. This factor level combination was defined as a generating design point since this combination was capable of estimating any one of the 16 uninvestigated interactions. In other words, using an additional design point which included the generating design point, an arbitrary uninvestigated interaction selected among the 16 interactions could be estimated. This generating design point could be found by inspecting the fourth factor level combination necessary to investigate each uninvestigated interaction. Table 22 shows the fourth factor level combination for each uninvestigated interaction. For example, the fourth factor level combination between MUS and DB was $(X_3, X_5) = (+1, -1) = (61.1 \text{ dBA, 2x6})$ since the combined data set provided the other three factor level combinations (i.e., $(X_3, X_5) = (-1.4142, +1), (-1, +1), (+1, +1)$ were included in the combined data set). Therefore, the factor level of the MUS factor in the generating design point was set at +1 coded value, and the factor level of the DB factor in the generating design point was set at -1 coded value. The factor levels of the other factors in the generating design point were selected in this fashion. The SR, IT, and SEX were not included in the generating design point, since the combined data set provided four different design points necessary to investigate the interaction effects between these and the other factors. In selecting additional design points, the experimenter must make sure that the additional design points cover all the factor levels defined in the generating design point since each uninvestigated interaction will be examined one at a time by using the additional design points selected.

Another computer run was performed to select the second design point necessary to investigate the interactions between AGE and the other variables. In running the computer program, the first design point selected was included in the combined data set to find the second design point because the first design point should affect the determinant values of the $X'X$ matrices. Table 27 shows the results of the computer run. As shown in Table 27, the candidate point for the uninvestigated interaction between X_3 and X_8 (i.e., between MUS and TV) had the maximum value among the other interactions. The factor level combination selected as the second design point was (SR, IT, MUS, AGE, DB, FB, WG, TV, NT, SEX) = (222 wpm, 3 secs, 61.1 dBA, Middle Age, 2x6, Available, Available, Betty, Two, Male). These two additional design points together satisfied the factor level combination defined in the generating design point. With the two design points selected, the 16 uninvestigated could be examined for their statistical significance ($\alpha=0.05$).

Before collecting data with the selected design points, another analysis was performed to check if the criteria for model item selection (i.e., the VIF and the eigen value ratio discussed in the previous section) were satisfied. Regression analyses were conducted using the SAS package (PROC REG with COLLINOINT and VIF options). The VIF values and the ratios were then examined to make sure that these additional design points satisfied the criteria (i.e., $VIF < 10$, and the ratio < 1000). If these criteria are not satisfied, it means that the effects of the uninvestigated interactions can not be properly examined due to serious multicollinearity among the model items. In this case, additional design points should be added until these criteria are satisfied. Table 28 through Table 31 show the results of this analysis for each dependent measure.

TABLE 27

Maximum $|X'X|$ Value for Each Uninvestigated Interaction (Second Design Point)

Uninvestigated Interaction	Max $ X'X $ Value
$X_3X_5^*$	0.2644×10^{30}
X_3X_6	0.9704×10^{30}
X_3X_7	0.2649×10^{30}
X_3X_8	$0.9879 \times 10^{30**}$
X_3X_9	0.2605×10^{30}
X_4X_5	0.9362×10^{29}
X_4X_6	0.5082×10^{30}
X_4X_7	0.9362×10^{29}
X_4X_8	0.5219×10^{30}
X_4X_9	0.9362×10^{29}
X_5X_8	0.5135×10^{30}
X_5X_9	0.1784×10^{30}
X_6X_8	0.8224×10^{30}
X_6X_9	0.5136×10^{30}
X_7X_8	0.5135×10^{30}
X_7X_9	0.1788×10^{30}

* X_3 =MUS, X_4 =AGE, X_5 =DB, X_6 =FB, X_7 =WG, X_8 =TV, X_9 =NT
 ** Maximum Value

As shown in Table 28, the VIF values of X_4X_5 and X_4X_7 for TST measure were 11.09 and 11.09 respectively, which were greater than the criterion (i.e., 10), although the ratios were less than 1000. Similarly, X_4X_5 and X_4X_7 for UAT and AKP had the VIF values greater than the criterion. Therefore, another computer run was performed to select the third design point. Six additional design points were selected in this fashion to satisfy these criteria. Table 32 shows the improvement of the VIF values as the six additional design points were added to the combined data set. At the end of the sixth design point, the VIF values satisfied the criterion. Table 33 presents the treatment conditions for the six additional design points. Data collection with the six additional design points (Experiment 4) were conducted. Experimental methods and procedures are described in the following section.

Experiment 4

This experiment was intended to obtain additional information necessary to bridge the individual data sets. The uninvestigated interactions could be examined by adding additional design points to the existing data sets. Data collection at the six additional design points were conducted to test for the significance of the 16 uninvestigated interactions.

The experimental methods and procedures were exactly the same as those described in the previous section except the background music level. The background music level was set at 61.1 dBA at the subject's ear (i.e., at +1 coded value). A Realistic STA-19 stereo receiver, a Realistic SCT-45 stereo dubbing cassette deck, and a pair of Realistic Minimus 7 speakers were

TABLE 28

Analysis of VIF and Eigen Value Ratio for TST (With Two Additional Design Points)

Uninvestigated Interaction	VIF Value	Max Eigen Value/ Min Eigen Value
X ₃ X ₅ *	5.57	2.43/0.08
X ₃ X ₆	1.53	2.43/0.14
X ₃ X ₇	5.57	2.43/0.08
X ₃ X ₈	1.58	2.59/0.14
X ₃ X ₉	1.58	2.59/0.14
X ₄ X ₅	11.09	2.43/0.04
X ₄ X ₆	2.15	2.43/0.14
X ₄ X ₇	11.09	2.43/0.04
X ₄ X ₈	2.03	2.68/0.14
X ₄ X ₉	2.03	2.69/0.14
X ₅ X ₈	1.99	2.57/0.14
X ₅ X ₉	1.99	2.57/0.14
X ₆ X ₈	1.24	2.43/0.14
X ₆ X ₉	1.24	2.43/0.14
X ₇ X ₈	1.99	2.57/0.14
X ₇ X ₉	1.99	2.57/0.14

* X₃=MUS, X₄=AGE, X₅=DB, X₆=FB, X₇=WG, X₈=TV, X₉=NT

TABLE 29
 Analysis of VIF and Eigen Value Ratio for UAT (With Two Additional
 Design Points)

Uninvestigated Interaction	VIF Value	Max Eigen Value Min Eigen Value
X ₃ X ₅ *	5.57	2.24/0.08
X ₃ X ₆	1.53	2.21/0.31
X ₃ X ₇	5.57	2.24/0.08
X ₃ X ₈	1.58	2.59/0.52
X ₃ X ₉	1.58	2.59/0.32
X ₄ X ₅	10.34	2.22/0.04
X ₄ X ₆	2.13	2.29/0.22
X ₄ X ₇	10.34	2.22/0.04
X ₄ X ₈	2.01	2.67/0.25
X ₄ X ₉	2.01	2.67/0.25
X ₅ X ₈	1.99	2.57/0.26
X ₅ X ₉	1.99	2.57/0.26
X ₆ X ₈	1.24	2.42/0.42
X ₆ X ₉	1.24	2.42/0.42
X ₇ X ₈	1.99	2.57/0.26
X ₇ X ₉	1.99	2.57/0.26
* X ₃ =MUS, X ₄ =AGE, X ₅ =DB, X ₆ =FB, X ₇ =WG, X ₈ =TV, X ₉ =NT		

TABLE 30

Analysis of VIF and Eigen Value Ratio for AKP (With Two Additional Design Points)

Uninvestigated Interaction	VIF Value	Max Eigen Value Min Eigen Value
X ₃ X ₅ *	6.60	2.28/0.07
X ₃ X ₆	1.64	2.21/0.30
X ₃ X ₇	6.60	2.28/0.07
X ₃ X ₈	1.63	2.60/0.31
X ₃ X ₉	6.60	2.67/0.07
X ₄ X ₅	10.94	2.22/0.04
X ₄ X ₆	2.18	2.30/0.22
X ₄ X ₇	10.94	2.22/0.04
X ₄ X ₈	2.12	2.67/0.23
X ₄ X ₉	10.94	2.72/0.04
X ₅ X ₈	2.02	2.52/0.26
X ₅ X ₉	6.21	2.56/0.10
X ₆ X ₈	1.25	2.37/0.41
X ₆ X ₉	2.00	2.41/0.31
X ₇ X ₈	2.02	2.52/0.26
X ₇ X ₉	6.21	2.56/0.10

* X₃=MUS, X₄=AGE, X₅=DB, X₆=FB, X₇=WG, X₈=TV, X₉=NT

TABLE 31
Analysis of VIF and Eigen Value Ratio for STA (With Two Additional Design Points)

Uninvestigated Interaction	VIF Value	Eigen Value Min Eigen Value
X ₃ X ₅ *	1.13	2.50/0.26
X ₃ X ₆	1.36	2.49/0.26
X ₃ X ₇	1.31	2.51/0.26
X ₃ X ₈	1.32	2.52/0.26
X ₃ X ₉	1.32	2.52/0.26
X ₄ X ₅	1.75	2.66/0.23
X ₄ X ₆	1.79	2.65/0.23
X ₄ X ₇	1.74	2.67/0.23
X ₄ X ₈	1.77	2.67/0.23
X ₄ X ₉	1.77	2.67/0.23
X ₅ X ₈	1.28	2.63/0.29
X ₅ X ₉	1.28	2.63/0.29
X ₆ X ₈	1.29	2.64/0.29
X ₆ X ₉	1.29	2.64/0.29
X ₇ X ₈	1.29	2.62/0.28
X ₇ X ₉	1.29	2.62/0.28
* X ₃ =MUS, X ₄ =AGE, X ₅ =DB, X ₆ =FB, X ₇ =WG, X ₈ =TV, X ₉ =NT		

TABLE 32

Analysis of VIF (With Six Additional Design Points)

Uninvestigated Interaction	VIF Value			
	TST	UAT	AKP	STA
X ₃ X ₅ *	3.09	3.09	3.80	1.30
X ₃ X ₆	1.31	1.31	1.53	1.26
X ₃ X ₇	3.09	3.09	3.80	1.28
X ₃ X ₈	1.56	1.56	1.58	1.24
X ₃ X ₉	1.56	1.56	3.26	1.24
X ₄ X ₅	5.47	5.46	6.06	1.67
X ₄ X ₆	1.78	1.77	1.86	1.59
X ₄ X ₇	5.52	5.52	5.98	1.66
X ₄ X ₈	1.75	1.74	1.84	1.56
X ₄ X ₉	1.75	1.75	4.51	1.56
X ₅ X ₈	1.66	1.66	1.71	1.24
X ₅ X ₉	1.66	1.66	3.61	1.24
X ₆ X ₈	1.26	1.26	1.26	1.31
X ₆ X ₉	1.26	1.26	1.80	1.31
X ₇ X ₈	1.65	1.65	1.71	1.24
X ₇ X ₉	1.65	1.65	3.61	1.24

* X₃=MUS, X₄=AGE, X₅=DB, X₆=FB, X₇=WG, X₈=TV, X₉=NT

TABLE 33
Treatment Conditions of Additional Design Points

Treatment	Factor					Level				
Condition	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀ *
1	-1	+1	+1	-1	-1	+1	+1	-1	+1	+1
2	+1	-1	+1	+1	-1	+1	+1	-1	+1	+1
3	+1	+1	+1	+1	+1	+1	+1	-1	+1	-1
4	-1	-1	+1	-1	-1	+1	-1	-1	+1	+1
5	+1	-1	+1	-1	+1	+1	+1	-1	+1	+1
6	-1	+1	+1	+1	-1	+1	-1	-1	+1	+1

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₆=FB, X₇=WG, X₈=TV, X₉=NT, X₁₀=SEX

connected to provide the background music. Background music was provided by playing a 45-minute audio cassette tape containing seven different songs recorded by the Muzak company. A Larson-Davis Precision Integrating Sound Level Meter (Model 800B) was used to set the background music level. The sound level was measured using the equivalent noise level (L_{eq}) methods as described in Williams (1978).

Results and Analysis

The results of Experiment 4 were added to the combined data set. Appendix B presents the integrated data set. The integrated data set (i.e., combined data set with the six additional observations) were then used to test the 16 uninvestigated interactions at $\alpha = 0.05$. Each uninvestigated interaction was added to the existing model items resulting from the sequential studies (see Table 23) and then a partial F-test for each uninvestigated interaction was performed. In other words, each uninvestigated interaction was added one at a time to the existing model items and then tested for its significance using a partial F-test. Table 34 shows the results of these analyses for each dependent measure. For the TST and STA measures, no interactions were found to be significant ($p > 0.05$). Therefore, only the existing model items resulting from the sequential studies were used to build empirical models for these measures.

On the other hand, the interaction effect between X_4 and X_7 (i.e., AGE x WG) for the AKP measure was found to be significant ($p < 0.05$). This new significant effect was then added to the existing model items to build an integrated empirical model for AKP measure. The interaction effects between X_4 and X_5 (i.e., AGE x DB), and X_4 and X_7 (i.e., AGE x WG) for the UAT

TABLE 34

Results of Partial F-Tests for Uninvestigated Interactions

Uninvestigated Interaction	p Value			
	TST	UAT	AKP	STA
X ₃ X ₅ *	0.3732	0.6962	0.5520	0.5199
X ₃ X ₆	0.1623	0.3900	0.5492	0.6657
X ₃ X ₇	0.2886	0.9797	0.9709	0.6540
X ₃ X ₈	0.9326	0.9722	0.9932	0.2812
X ₃ X ₉	0.7002	0.6187	0.8573	0.1936
X ₄ X ₅	0.4443	0.0048**	0.5547	0.8355
X ₄ X ₆	0.2788	0.0566	0.8553	0.9772
X ₄ X ₇	0.6-79	0.0304**	0.0389**	0.6869
X ₄ X ₈	0.8410	0.2781	0.7066	0.1484
X ₄ X ₉	0.9476	0.1342	0.6894	0.0979
X ₅ X ₈	0.6275	0.6660	0.5964	0.6854
X ₅ X ₉	0.8287	0.9857	0.5520	0.5503
X ₆ X ₈	0.7130	0.8333	0.6971	0.5843
X ₆ X ₉	0.5454	0.5652	0.5547	0.4451
X ₇ X ₈	0.7207	0.4771	0.9239	0.5800
X ₇ X ₉	0.9292	0.7665	0.9709	0.4556

* X₃=MUS, X₄=AGE, X₅=DB, X₆=FB, X₇=WG, X₈=TV, X₉=NT** Significant at $\alpha=0.05$

measure were found to be significant ($p < 0.05$). These two effects were included in building an integrated empirical model for UAT measure.

Before building the empirical model for the UAT measure, an analysis was performed to make sure that these two interactions could be included in the model at the same time since the test for significance was conducted one at a time. That is, if there had been more than six uninvestigated interactions found to be significant from the individual F-tests, then additional experimental runs to the six design points selected should have been conducted since the six additional design points were capable of estimating at most six different interactions. Since there were only two additional significant interactions for the UAT, which resulted from the partial F-tests, additional data collection was not necessary.

In the next chapter, the procedures for building empirical models are presented in detail.

CHAPTER 6

INTEGRATED EMPIRICAL MODELS

Based on the results of the data bridging described in the previous chapter, individual empirical models for each performance measure were developed using polynomial regression. Table 35 through Table 38 show the results of these regression analyses. From the Table 35, the empirical model for the TST measure can be expressed using polynomials as shown in Eq. (3) since there were no additional interactions which were found to be significant using partial F-tests. In other word, the empirical model for the TST measure consisted of only the effects which were found to be significant from the previous three studies (i.e., Experiment 1, 2, and 3). Likewise, the STA measure is expressed in Eq. (6) which contains the significant effects resulting from the previous studies since there were no additional interactions found to be significant during the data bridging procedures. On the other hand, the UAT measure shown in Eq. (4) has two interactions in addition to the significant effects found in the previous studies (i.e., X_4X_5 and X_4X_7). The AKP measure shown in Eq. (5) has one additional interaction (i.e., X_4X_7).

$$\begin{aligned} \text{TST} = & 52.81 - 1.20X_1 + 17.27X_2 + 2.77X_3 + 2.15X_4 + 8.27X_5 - 4.47X_7 \\ & + 1.66X_2X_3 + 4.49X_2X_5 - 1.55X_2X_7 + 2.83X_5X_7. \end{aligned} \quad \text{Eq. (3)}$$

TABLE 35

Results of Regression Analysis for TST

Variable	Parameter Estimate	Standard Error	t for H ₀ : Parameter=0	p Value
Intercept	52.81	1.40	37.70	0.0001
X ₁ *	-1.20	0.67	-1.79	0.0762
X ₂	17.27	1.26	13.76	0.0001
X ₃	2.77	0.99	2.79	0.0061
X ₄	2.15	1.10	1.96	0.0526
X ₅	8.27	1.10	7.51	0.0001
X ₇	-4.47	1.10	-4.06	0.0001
X ₂ X ₃	1.66	0.86	1.93	0.0551
X ₂ X ₅	4.49	0.80	5.62	0.0001
X ₂ X ₇	-1.55	0.80	-1.94	0.0543
X ₅ X ₇	2.83	1.10	2.57	0.0113

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG

TABLE 36

Results of Regression Analysis for UAT

Variable	Parameter Estimate	Standard Error	t for H ₀ : Parameter=0	p Value
Intercept	9.15	2.15	4.26	0.0001
X ₁ *	0.65	0.71	0.92	0.3615
X ₂	1.74	0.84	2.08	0.0392
X ₃	4.05	1.06	3.80	0.0002
X ₄	-4.67	2.33	-2.00	0.0472
X ₅	-0.69	2.16	-0.32	0.7491
X ₇	-6.03	2.17	-2.78	0.0062
X ₂ X ₅	-1.58	0.85	-1.86	0.0654
X ₂ X ₇	-1.28	0.85	-1.51	0.1340
X ₅ X ₇	2.30	1.16	1.98	0.0496
X ₄ X ₅	4.46	2.14	2.08	0.0396
X ₄ X ₇	-2.15	2.16	-1.00	0.3198

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG

TABLE 36

Results of Regression Analysis for UAT

Variable	Parameter Estimate	Standard Error	t for H ₀ : Parameter=0	p Value
Intercept	9.15	2.15	4.26	0.0001
X ₁ *	0.65	0.71	0.92	0.3615
X ₂	1.74	0.84	2.08	0.0392
X ₃	4.05	1.06	3.80	0.0002
X ₄	-4.67	2.33	-2.00	0.0472
X ₅	-0.69	2.16	-0.32	0.7491
X ₇	-6.03	2.17	-2.78	0.0062
X ₂ X ₅	-1.58	0.85	-1.86	0.0654
X ₂ X ₇	-1.28	0.85	-1.51	0.1340
X ₅ X ₇	2.30	1.16	1.98	0.0496
X ₄ X ₅	4.46	2.14	2.08	0.0396
X ₄ X ₇	-2.15	2.16	-1.00	0.3198

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG

TABLE 37

Results of Regression Analysis for AKP

Variable	Parameter Estimate	Standard Error	t for H ₀ : Parameter=0	p Value
Intercept	1.02	0.10	10.25	0.0001
X ₃ *	0.13	0.05	2.56	0.0117
X ₄	0.28	0.10	2.69	0.0080
X ₅	-0.34	0.06	-5.95	0.0001
X ₇	-0.11	0.10	-1.05	0.2935
X ₉	0.12	0.05	2.24	0.0265
X ₂ X ₇	-0.07	0.04	-1.96	0.0520
X ₅ X ₇	0.23	0.06	4.03	0.0001
X ₄ X ₇	0.21	0.10	2.09	0.0389

* X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG, X₉=NT

TABLE 38

Results of Regression Analysis for STA

Variable	Parameter Estimate	Standard Error	t for H ₀ : Parameter=0	p Value
Intercept	3.16	0.08	38.20	0.0001
X ₁ *	-0.15	0.02	-5.94	0.0001
X ₃	-0.40	0.04	-9.47	0.0001
X ₄	-0.15	0.04	-3.44	0.0008
X ₂ X ₁₀	-0.02	0.03	-0.87	0.3871
X ₃ ²	-0.21	0.07	-3.16	0.0019

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₁₀=SEX

$$\begin{aligned} \text{UAT} = & 9.14 + 0.65X_1 + 1.74X_2 + 4.05X_3 - 4.67X_4 - 0.69X_5 - 6.03X_7 \\ & - 1.58X_2X_5 - 1.28X_2X_7 + 2.30X_5X_7 + 4.46X_4X_5 - 2.15X_4X_7. \end{aligned} \quad \text{Eq. (4)}$$

$$\begin{aligned} \text{AKP} = & 1.02 + 0.13X_3 + 0.28X_4 - 0.34X_5 - 0.11X_7 + 0.12X_9 - 0.07X_2X_7 \\ & + 0.23X_5X_7 + 0.21X_4X_7. \end{aligned} \quad \text{Eq. (5)}$$

$$\text{STA} = 3.16 - 0.15X_1 - 0.40X_3 - 0.15X_4 - 0.02X_2X_{10} - 0.21X_3^2. \quad \text{Eq. (6)}$$

The independent variables in these four equations used coded values. For example, the value of the continuous variables (i.e., X_1 , X_2 , and X_3) ranged from -1.414 to +1.414 as defined in Experiment 1 (Wu, 1989). For a discrete variable, -1 and +1 coded values were used to build these equations. The TST measure describes the total search time per trial, and the UAT measure explains the user added time beyond the error-free search task per trial. The AKP measure shows the average number of user added keypresses per trial and the STA describes the average number of transcription scores per trial.

Model Selection

There can be serious multicollinearity when many uninvestigated interactions are added to the existing model items at the same time. In other words, each significant uninvestigated interaction can be seriously confounded with the existing model items when many additional interactions are

simultaneously added to the existing model items. The effects of multicollinearity include (1) large variances on coefficients; (2) poor power of partial-F tests; (3) incorrect signs on coefficients; and (4) poor prediction performance (Myers, 1990).

After determining significant uninvestigated interactions (see Chapter 5), these significant interactions were added to the existing model items at the same time. However, the simultaneous addition of all significant interactions can cause multicollinearity problems since each uninvestigated interaction was tested one by one under the presence of the existing model items during the data bridging procedures. For example, the interaction effect between X_4 and X_7 in the Eq. (4) was found to be insignificant ($p > 0.05$) when it was added to the existing model items together with the interaction X_4X_5 . The interaction X_4X_7 , however, was found to be significant ($p < 0.05$) when it was tested individually against the existing model items. In other words, the role of the interaction X_4X_7 was changed when it was added together with the interaction X_4X_5 to the existing model items. This effect could have been more serious with the addition of more than two significant interactions to the existing model items. To avoid these problems, several decision rules were developed as a means of obtaining better empirical models in terms of the prediction performance and variances of coefficients of the empirical models.

Rule (1): Significant model items resulting from the previous studies (i.e., Experiment 1, 2, and 3) should always remain in the model.

These model items are considered more important than additional interactions because of the following two reasons: first, these model items were investigated unconfounded with any other main

effects and two-factor interactions; and second, at the early stage of the sequential research strategy, all possible interactions which might be important for describing the performance measures could be planned to be investigated orthogonally (e.g., this plan could be developed by grouping independent variables in such a way that these important interactions should be investigated unconfounded with other main effects and two-factor interactions.). This rule has been applied to the data bridging procedures. That is, each uninvestigated interaction was tested under the presence of the existing model items.

Rule (2): If no more than one additional interaction is found to be significant during the data bridging procedures, then additional analysis to obtain a better empirical model is not necessary. In other words, the final model includes the existing model items and one additional interaction (if any). This rule is based on the fact that the partial F-tests during the data bridging procedures were conducted under the presence of the existing model items. Therefore, multicollinearity criteria (i.e., VIF and eigen value ratio) were satisfied to test this uninvestigated interaction.

Rule (3): If there are more than one additional interactions which are added to the existing model items, a model selection procedure is necessary to obtain a better empirical model.

Using these decision rules, the empirical models for the TST, AKP, and STA measures were not subject to further analyses since these models satisfied the decision Rule (2). However, the empirical model for the UAT measure

shown in Eq. (4) did not satisfy the decision Rule (2). Therefore, the UAT model was analyzed using several model selection criteria such as model mean square errors, C_p statistic, and PRESS statistic described by Myers (1990). The model mean square error describes the degree of data variations unexplained by a regression equation. The C_p statistic explains the degree of prediction variances and biases of the model. Finally, the PRESS statistic provides a separate measure of the stability of the regression by setting each observation in calculating residual errors of the model. Smaller values of these statistics provide a better regression equation in terms of prediction variance and variances of coefficients.

Selection Procedures

Model selection procedures began with selecting candidate models for examination. Based on the Rule (1) described in the previous section, the existing model items were included in all candidate models. Since there were two additional interactions (i.e., X_4X_5 and X_4X_7) for the STA measure, candidates were divided into three models (1) Model 1: the existing model items with X_4X_5 ; (2) Model 2: the existing model items with X_4X_7 ; and (3) Model 3: the existing model items with X_4X_5 and X_4X_7 . The values of the three statistics (i.e., mean square errors, C_p statistics, and PRESS statistics) for each model were calculated using a SAS program (PROC REG and PROC RSQUARE). In addition to these statistics, average confidence intervals, DFFITS, and R-student values (Myers, 1990) were used in the selection procedures to diagnose the candidates. Table 39 shows the results of this analysis. As shown in Table 39, all models had the same number of R-student and DFFITS values that did not satisfy the criteria. Model 1 (i.e., the existing model items

TABLE 39
Results of Model Selection for UAT

Candidate Model	Cp	PRESS	MSE	Average C.I.	Number of R-students > 3.6	Number of DFFITS >2.0
Model 1*	11.0	15860	95.488	9.60	3	1
Model 2	14.3	16700	97.741	9.72	3	1
Model 3	12.0	15658	95.491	9.83	3	1

* Model 1: $X_1, X_2, X_3, X_4, X_5, X_7, X_2X_5, X_2X_7, X_5X_7, X_4X_5$
Model 2: $X_1, X_2, X_3, X_4, X_5, X_7, X_2X_5, X_2X_7, X_5X_7, X_4X_7$
Model 3: $X_1, X_2, X_3, X_4, X_5, X_7, X_2X_5, X_2X_7, X_5X_7, X_4X_5, X_4X_7$

with X_4X_5) had the smallest value of C_p , mean square error, and average confidence interval, although the value of PRESS of the Model 1 was greater than that of the Model 3. These results showed that it was not necessary to include the interaction X_4X_7 when the model had the interaction X_4X_5 . Based on these statistics, Model 1 was selected as the final model for the UAT measure. As a result, a new regression analysis was conducted to obtain new coefficients. Table 40 shows the results of this regression analysis. The variances of coefficients shown in Table 40 were smaller than those shown in Table 36 by eliminating the interaction X_4X_7 . The new model for the UAT measure was then redeveloped using the results shown in Table 40.

$$\begin{aligned} \text{UAT} = & 9.96 + 0.57X_1 + 1.76X_2 + 3.95X_3 - 3.61X_4 + 0.18X_5 - 4.20X_7 \\ & - 1.72X_2X_5 - 1.38X_2X_7 + 2.29X_5X_7 + 5.45X_4X_5. \end{aligned} \qquad \text{Eq. (4)*}$$

By comparing the Eq. (4) and Eq. (4)*, there were several differences between them. First, the Eq. (4) did not have the interaction of X_4X_7 . Second, the sign of the coefficient of X_5 was changed from a negative to a positive. Finally, the values of the coefficients were changed. The new model had better performances in terms of prediction variances and variances of coefficients in describing the functional relationships of the independent variables.

Integrated Empirical Models

After completing model selection procedures, integrated empirical models for each performance measure were obtained using coded values. Eq.

TABLE 40

Results of Regression Analysis for UAT Using a New Model

Variable	Parameter Estimate	Standard Error	t for H ₀ : Parameter=0	p Value
Intercept	9.96	1.99	5.01	0.0001
X ₁ *	0.57	0.71	0.81	0.4198
X ₂	1.76	0.84	2.11	0.0369
X ₃	3.95	1.06	3.72	0.0003
X ₄	-3.61	2.08	-1.74	0.0844
X ₅	0.18	1.97	0.09	0.9264
X ₇	-4.20	1.16	-3.61	0.0004
X ₂ X ₅	-1.72	0.84	-2.05	0.0418
X ₂ X ₇	-1.38	0.84	-1.64	0.1042
X ₅ X ₇	2.28	1.16	1.96	0.0515
X ₄ X ₅	5.45	1.90	2.87	0.0048

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG

(3), (4)*, (5), and (6) describe the functional relationships among independent variables for the TST, UAT, AKP, and STA measures, respectively. These equations are re-presented below;

$$\begin{aligned} \text{TST} = & 52.81 - 1.20X_1 + 17.27X_2 + 2.77X_3 + 2.15X_4 + 8.27X_5 - 4.47X_7 \\ & + 1.66X_2X_3 + 4.49X_2X_5 - 1.55X_2X_7 + 2.83X_5X_7. \end{aligned} \quad \text{Eq. (3)}$$

$$\begin{aligned} \text{UAT} = & 9.96 + 0.57X_1 + 1.76X_2 + 3.95X_3 - 3.61X_4 + 0.18X_5 - 4.20X_7 \\ & - 1.72X_2X_5 - 1.38X_2X_7 + 2.29X_5X_7 + 5.45X_4X_5. \end{aligned} \quad \text{Eq.(4)*}$$

$$\begin{aligned} \text{AKP} = & 1.02 + 0.13X_3 + 0.28X_4 - 0.34X_5 - 0.11X_7 + 0.12X_9 - 0.07X_2X_7 \\ & + 0.23X_5X_7 + 0.21X_4X_7. \end{aligned} \quad \text{Eq. (5)}$$

$$\text{STA} = 3.16 - 0.15X_1 - 0.40X_3 - 0.15X_4 - 0.02X_2X_{10} - 0.21X_3^2. \quad \text{Eq. (6)}$$

In these equations, an X_i shows the main effect of X_i variable; an X_iX_j shows the linear by linear interaction effect between X_i and X_j ; and an X_i^2 shows the pure quadratic effect of X_i variable. Each coefficient of the model items represents a rate of change when each model item changes by one unit of coded values. The discrete variables (i.e., X_4 through X_{10}) can have either a -1 or a +1 coded value in these equations since they have only two levels (e.g., Available or Not Available); while the continuous variables can have any value ranging from -1.4142 to +1.4142 coded values. Table 41 shows natural values of independent variables corresponding to coded values. For example, -1.4142 coded value of X_1 (i.e., SR) corresponds to 120 (wpm); while +1.4142

coded value corresponds to 240 (wpm) in natural unit. In converting coded values of continuous variables, $\text{Min}_{(\text{natural})} + ((\text{Max}_{(\text{natural})} - \text{Min}_{(\text{natural})}) \times (X_i + 1.4142))/2.8284$ was used, where $\text{Min}_{(\text{natural})}$ is the lower bound of the variable X_i in natural unit, $\text{Max}_{(\text{natural})}$ is the upper bound of the variable X_i in natural unit, and X_i is the variable in coded value. For example, to obtain the natural value of +1 coded value of X_1 (i.e., SR), 120 (wpm) for $\text{Min}_{(\text{natural})}$ and +1 for X_i can be inserted into the formula and then the formula gives approximately 222 (wpm) of natural value of +1 coded value. For an example of discrete variables, -1 coded value of X_4 (i.e., AGE) corresponds to college age subjects; while +1 coded value corresponds to middle age subjects.

Interpretation of Empirical Models

The empirical models for the TST, UAT, AKP, and STA measures had R^2 values of 0.8702, 0.3199, 0.4199, and 0.5913, respectively. From the functional relationships, it can be readily observed what model items are involved in describing the data variation for each measure. In addition, the magnitude of each coefficient shows the relative role of each regressor in describing each dependent measure when each regressor changes by one coded value. For example, six main effects (i.e., X_1 , X_2 , X_3 , X_4 , X_5 , and X_7) and four two-factor interactions (i.e., X_2X_3 , X_2X_5 , X_2X_7 , and X_5X_7) described the system performance in terms of TST measure. In addition, it can be easily found what regressors have effects on what dependent measure. Table 42 summarizes the model items used for developing the empirical model for each dependent measure. As shown in Table 42, X_1 (i.e., SR) main effect is included in the models for the TST, UAT, and STA measures. However, SR has no effect on the AKP measure.

TABLE 41
Conversion of Variables From Coded Values To Natural Values

Variable	Coded Value					
	-1.4142	-1.0	0.0	+1.0	+1.4142	
X ₁ (SR)	120	138	180	222	240	(wpm)
X ₂ (IT)	2	3	6	8	10	(secs)
X ₃ (MUS)	36.1	40.5	50.8	61.1	65.4	(dBA)
X ₄ (AGE)*	CA		MA			
X ₅ (DB)	2x6		8x2			
X ₆ (FB)	No		Yes			
X ₇ (WG)	No		Yes			
X ₈ (TV)	Betty		Paul			
X ₉ (NT)	One		Two			
X ₁₀ (SEX)	Female		Male			

* CA = College Age Subject (18 - 30 years),
 MA = Middle Age Subject (36 - 60 years)

TABLE 42

Model Items Used For Developing Empirical models

Model Item	Dependent Measure			
	TST	UAT	AKP	STA
X ₁ *	Y**	Y	-	Y
X ₂	Y	Y	-	-
X ₃	Y	Y	Y	Y
X ₄	Y	Y	Y	Y
X ₅	Y	Y	Y	-
X ₇	Y	Y	Y	-
X ₉	-	-	Y	-
X ₂ X ₃	Y	-	-	-
X ₂ X ₅	Y	Y	-	-
X ₂ X ₇	Y	Y	Y	-
X ₂ X ₁₀	-	-	-	Y
X ₄ X ₅	-	Y	-	-
X ₄ X ₇	-	-	Y	-
X ₅ X ₇	Y	Y	Y	-
X ₃ ²	-	-	-	Y

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG, X₉=NT, X₁₀=SEX

** Y: Used, -: Not Used

In the TST measure, one unit increase of X_2 main effect (i.e., 2.8 second increase of input timeout) is equivalent to approximately 6 unit increase of X_3 main effect (i.e., 62.2 dBA increase of background music level) since the ratio of the coefficients of these two variables is approximately 6 (i.e., $17.27/2.77$). From the functional relationships, the researcher can determine what variables are important in designing an interface and how much each variable affects the performance at the interface. If the researcher wishes to predict human performance with new combinations of independent variables, the functional relationships can provide predicted values and prediction variances by using SAS program (PROC REG). For example, at a new combination of independent variable, e.g., $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, \text{ and } X_{10}) = (0.5, 0.5, 0.5, -1, -1, -1, -1, -1, -1, -1) = (201 \text{ wpm}, 7.4 \text{ secs}, 55.9 \text{ dBA}, \text{ College Age}, 2 \times 6, \text{ No Feedback}, \text{ No Wallet Guide}, \text{ Betty}, \text{ One}, \text{ Female})$, the predicted values with standard errors of prediction for the TST, UAT, AKP, and STA are 58.05 secs with 3.19, 30.01 secs with 3.38, 1.61 with 0.16, and 3.00 with 0.10, respectively.

Further Analysis

Further analyses were performed to identify possible outliers and highly influential points in the integrated data set. If any outliers or influential points exist, these observations should be checked since they may be a result of a circumstance external to the system under study, e.g., key punch errors, faulty observation, etc. (Myers, 1990). The R-student statistic for outlier detection and DFFITS for detection of influential points were used in the analyses. The value of 3.61 was used for an outlier detection criterion ($\alpha=0.05$ approximately) and the value of 2 was used as a yardstick for DFFITS ($\alpha=0.05$).

For the TST measure, two observations (observation number 22 and 41) resulting from Experiment 1 were found to be possible outliers (see Appendix B for the observation number). The factor level combinations of these observations were $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, \text{ and } X_{10}) = (0, 0, 0, 1, 1, -1, -1, 1, -1, 1) = (180 \text{ wpm}, 6 \text{ secs}, 50.8 \text{ dBA}, \text{ Middle Age}, 8 \times 2, \text{ No Feedback}, \text{ No Wallet Guide}, \text{ Paul}, \text{ One}, \text{ Male})$ for the observation 22 and $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, \text{ and } X_{10}) = (1, 1, 1, 1, 1, -1, -1, 1, -1, 1) = (222 \text{ wpm}, 9 \text{ secs}, 61.1 \text{ dBA}, \text{ Middle Age}, 8 \times 2, \text{ No Feedback}, \text{ No Wallet Guide}, \text{ Paul}, \text{ One}, \text{ Male})$ for the observation 41. The responses at these two points were 99.3 secs and 161.2 secs, respectively. However, the DFFITS value only at the observation number 41 was found to be significant. Therefore, the observation number 22 did not appear to have any influence on the predicted values of the TST measure. The same results were obtained for the UAT measure. The responses at these observations were 48.2 secs and 90.9 secs, respectively.

On the other hand, the regression equation of the AKP measure had another observation (observation number 51) resulting from Experiment 2 in addition to the observation number 41 resulting from Experiment 1, which were identified as possible outliers. The factor level combination of the observation number 51 was $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, \text{ and } X_{10}) = (1.4142, 1.4142, -1.4142, -1, -1, 1, -1, 1, -1, -1) = (240 \text{ wpm}, 9 \text{ secs}, 36.1 \text{ dBA}, \text{ College Age}, 2 \times 6, \text{ Feedback}, \text{ No Wallet Guide}, \text{ Paul}, \text{ One}, \text{ Female})$. The response at these observations were 3.0 for the observation number 41, and 3.9 for the observation 51. However, there were no significant DFFITS values. These results showed that although the values of R-student statistics were found to be

significant, these two observations did not have any significant effects on the predicted responses.

For the UAT measure, observation number 148 resulting from Experiment 4 was found to be a possible outlier. The factor level combination at this observation was $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, \text{ and } X_{10}) = (1, -1, 1, 1, -1, 1, 1, -1, 1, 1) = (222 \text{ wpm}, 3 \text{ secs}, 61.1 \text{ dBA}, \text{ Middle Age}, 2 \times 6, \text{ Feedback}, \text{ Wallet Guide}, \text{ Betty}, \text{ Two}, \text{ Male})$. The response at this observation was 1.09. However, the DFFITS value was not significant.

In summary, there were four possible outliers in the integrated data set (i.e., observation number 22, 41, 51, and 148) and one influential point (i.e., observation number 41). It may be necessary to eliminate these observations if these data do not truly describe the nature of human performance because of possible external happenings (Myers, 1990). These data were checked against any experimental errors, but there was no evidence of happenings external to the system under study. Therefore, these data remained in the integrated data set.

CHAPTER 7

OPTIMIZATION

After building the empirical models, the models were analyzed to determine a configuration providing the maximum performance in the system. This procedure determined what values of independent variables maximized the performance of the system. A variety of approaches have been suggested for determining the point of maximum performance given a functional relationship between independent and dependent variables. These approaches include the canonical analysis and ridge analysis described by Box and Draper (1987) and Myers (1976), and other optimization techniques used in nonlinear programming (Avriel, 1976; Bazaraa and Shetty, 1979). However, since many independent variables used in the telephone information study were discrete (especially dichotomous), these techniques could not be directly applied to optimizing the integrated models which included discrete variables. An alternative approach to optimizing the performances was to use a separate optimization for each combination of discrete variables. In other words, the integrated model were separately analyzed for each combination of discrete variables and then compared with each other to obtain the treatment level which provides the maximum value of performance.

Although the canonical analysis and ridge analysis are capable of providing near optimum solutions in addition to an optimum solution for a

general second-order model, computer packages for these analyses were not available for analyzing the empirical models presented in the previous chapter. As an alternative, an IMSL computer package (BCPOL subroutine) was used to determine optimum solutions for the integrated empirical models. The BCPOL subroutine uses the complex method (Gill, Murray, and Wright, 1981) to find an optimum point of a function of n variables. This method is an appropriate method to find an optimum solution of a second-order function given a simple bound constraint on each independent variable. The method is based on function comparison. It starts with $2n$ points around an initial starting point (which must be supplied by the user) in the n -dimensional space generated by the n independent variables and then computes the function values at these points. At each iteration, a new point is generated to replace the worst point, which has the largest value among these $2n$ points (or the smallest value for a maximization problem). The new point is constructed using a specific algorithm in such a way that its function value is improved. The subroutine ends when a given tolerance criterion is satisfied. The tolerance is defined as the difference between the best function value and the worst function value. A 1.0×10^{-5} tolerance value was used in running the BCPOL subroutine.

Optimum Configurations

Optimum Configuration for TST Measure

The Eq. (3) consisted of three continuous variables (i.e., X_1 , X_2 , and X_3) and three discrete variables (i.e., X_4 , X_5 , and X_7). Therefore, eight different equations were developed to separate the Eq. (3) according to each

combination of these three discrete variables since each discrete variable had only two different levels. In other words, for a specific factor level combination of the three discrete variables, a new equation was developed by inserting each factor level of the discrete variables into the Eq. (3). For example, the factor level combination of $(X_4, X_5, X_6) = (-1, -1, -1)$ had a new equation of $TST = 49.69 - 1.20X_1 + 14.33X_2 + 2.77X_3 + 1.66X_2X_3$ as shown in the first row of Table 43.

Given the eight different functions, computer runs using the BC POL subroutine were conducted to find an optimum solution for each function. The constraint of the continuous variables (i.e., X_1, X_2 , and X_3) used ranged from -1.4142 to +1.4142 in coded value. The initial starting point used was the center point of the three continuous variables (i.e., $(X_1, X_2, X_3) = (0, 0, 0)$). Table 44 shows the results of these computer runs.

As shown in Table 44, all optimum solutions were obtained at the same design point in terms of the three continuous variables, i.e., at $(X_1, X_2, X_3) = (1.4142, -1.4142, -1.4142)$. The minimum TST value among the eight different functions was 16.9 seconds. The specific factor level combination which generated this minimum value was $(X_1, X_2, X_3, X_4, X_5, X_7) = (1.4142, -1.4142, -1.4142, -1, -1, 1) = (240 \text{ wpm}, 2 \text{ secs}, 36.1 \text{ dBA}, \text{College Age}, 2 \times 6, \text{Wallet Guide})$. In other word, when a college age user uses the telephone information system with the fastest speech rate, smallest input timeout, no background music, 2x6 database organization, and a wallet guide, the total search time is minimum.

TABLE 43
Eight Equations of TST Measure for Different Combinations of Discrete Factors

Equation	X ₄ *	X ₅	X ₇
49.69-1.20X ₁ +14.33X ₂ +2.77X ₃ +1.66X ₂ X ₃	-1	-1	-1
35.09-1.20X ₁ +11.23X ₂ +2.77X ₃ +1.66X ₂ X ₃	-1	-1	+1
60.57-1.20X ₁ +23.31X ₂ +2.77X ₃ +1.66X ₂ X ₃	-1	+1	-1
53.99-1.20X ₁ +14.33X ₂ +2.77X ₃ +1.66X ₂ X ₃	+1	-1	-1
64.87-1.20X ₁ +23.31X ₂ +2.77X ₃ +1.66X ₂ X ₃	+1	+1	-1
39.39-1.20X ₁ +11.23X ₂ +2.77X ₃ +1.66X ₂ X ₃	+1	-1	+1
57.29-1.20X ₁ +20.21X ₂ +2.77X ₃ +1.66X ₂ X ₃	-1	+1	+1
61.59-1.20X ₁ +20.21X ₂ +2.77X ₃ +1.66X ₂ X ₃	+1	+1	+1

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG

TABLE 44

Optimum Configurations for TST Measure

X_1	X_2	X_3	X_4	X_5	X_7	Optimum (secs)
1.4142	-1.4142	-1.4142	-1	-1	-1	27.1
1.4142	-1.4142	-1.4142	-1	-1	+1	16.9
1.4142	-1.4142	-1.4142	-1	+1	-1	25.3
1.4142	-1.4142	-1.4142	+1	-1	-1	31.4
1.4142	-1.4142	-1.4142	+1	+1	-1	29.6
1.4142	-1.4142	-1.4142	+1	-1	+1	21.2
1.4142	-1.4142	-1.4142	-1	+1	+1	26.4
1.4142	-1.4142	-1.4142	+1	+1	+1	30.7

* X_1 =SR, X_2 =IT, X_3 =MUS, X_4 =AGE, X_5 =DB, X_7 =WG

Optimum Configuration for UAT measure

Table 45 shows the eight separate equations by using different combinations of the discrete factor levels shown in Eq. (4)* since there were three discrete variables (i.e., X_4 , X_5 , and X_7). As shown in Table 45, all of these equations are in the form of a first-order empirical model in terms of the three continuous variables (i.e., X_1 , X_2 , and X_3). The optimum values can be easily obtained by assigning boundary points to each variable depending on its sign of the coefficient. For example, to minimize the UAT, the lower bound value of a variable can be used if the sign of the coefficient of a variable is positive (or upper bound value for a negative coefficient). In this fashion, the optimum values were obtained for the eight equations. Table 46 shows the optimum values for the eight equations.

As shown in Table 46, the minimum UAT value among the eight different functions was -15.1 seconds. This value did not show any practical implications because UAT values should be equal to or greater than zero (Sometimes, a subject can conduct the search task faster than an average ideal user. However, the UAT values are not expected to be far apart from zero). This nonsense value might result from the low R^2 value (0.3199) of the UAT equation and large prediction variances. For example, the standard error of prediction at the optimum point was 7.9 seconds. Although the optimum value was not practical, the factor level combination which provided the optimum performance might be useful to find the optimal configuration of independent variables. The specific factor level combination which generated this minimum UAT was $(X_1, X_2, X_3, X_4, X_5, X_7) = (-1.4142, -1.4142, -1.4142, 1, -1, 1) = (120 \text{ wpm}, 2 \text{ secs}, 36.1 \text{ dBA}, \text{Middle Age}, 2 \times 6, \text{Wallet Guide})$. In other

TABLE 45

Eight Equations of UAT Measure for Different Combinations of Discrete Factors

Equation	X ₄ *	X ₅	X ₇
25.32 +0.57X ₁ +4.86X ₂ +3.95X ₃	-1	-1	-1
12.36 +0.57X ₁ +2.10X ₂ +3.95X ₃	-1	-1	+1
10.22 +0.57X ₁ +1.42X ₂ +3.95X ₃	-1	+1	-1
7.20 +0.57X ₁ +4.86X ₂ +3.95X ₃	+1	-1	-1
13.90 +0.57X ₁ +1.42X ₂ +3.95X ₃	+1	+1	-1
-5.76 +0.57X ₁ +2.10X ₂ +3.95X ₃	+1	-1	+1
6.38 +0.57X ₁ -1.34X ₂ +3.95X ₃	-1	+1	+1
10.06 +0.57X ₁ -1.34X ₂ +3.95X ₃	+1	+1	+1

* X₁ =SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG

TABLE 46

Optimum Configurations for UAT Measure

X_1	X_2	X_3	X_4	X_5	X_7	Optimum (secs)
-1.4142	-1.4142	-1.4142	-1	-1	-1	12.1
-1.4142	-1.4142	-1.4142	-1	-1	+1	3.0
-1.4142	-1.4142	-1.4142	-1	+1	-1	1.8
-1.4142	-1.4142	-1.4142	+1	-1	-1	-6.1
-1.4142	-1.4142	-1.4142	+1	+1	-1	5.5
-1.4142	-1.4142	-1.4142	+1	-1	+1	-15.1
-1.4142	+1.4142	-1.4142	-1	+1	+1	1.9
-1.4142	+1.4142	-1.4142	+1	+1	+1	5.6

* X_1 =SR, X_2 =IT, X_3 =MUS, X_4 =AGE, X_5 =DB, X_7 =WG

words, when a middle age user uses the telephone information system with the slowest speech rate, smallest input time out, no background music, 2x6 database organization, and a wallet guide, the user added time is minimum. It is interesting to note that middle age subjects had a smaller optimum UAT than college age subjects at the optimum configuration.

Optimum Configuration for AKP measure

Table 47 shows the 16 separate equations by using different combinations of the discrete factor levels in Eq. (5) since there were four discrete variables (i.e., X_4 , X_5 , X_7 , and X_9). As shown in Table 47, all these equations are in the form of a first-order empirical model in terms of the three continuous variables (i.e., X_1 , X_2 , and X_3). The optimum values, therefore, were obtained by using the methods described in the previous section. Table 48 shows the optimum solutions for the 16 equations.

As shown in Table 48, the minimum AKP value among the 16 different functions was -0.1. This value did not show any practical implications because AKP values should be equal to or greater than zero. This nonsense value might result from the low R^2 value (0.4199) of the AKP equation and large prediction variances. For example, the standard error of prediction at the optimum point was 0.14. Although the optimum value was not practical, the factor level combination which provided the optimum performance might be useful to find the optimal configuration of independent variables. The specific factor level combination which generated this minimum AKP was $(X_2, X_3, X_4, X_5, X_7, X_9) = (1.4142, -1.4142, -1, 1, 1, -1) = (10 \text{ secs}, 36.1 \text{ dBA}, \text{College Age}, 8 \times 2, \text{Wallet Guide}, \text{One})$. In other words, when a college age user performs single target

TABLE 47

Sixteen Equations of AKP Measure for Different Combinations of Discrete Factors

Equation	X_4^*	X_5	X_7	X_9
$1.51 + 0.07X_2 + 0.13X_3$	-1	-1	-1	-1
$1.75 + 0.07X_2 + 0.13X_3$	-1	-1	-1	+1
$0.41 - 0.07X_2 + 0.13X_3$	-1	-1	+1	-1
$0.37 + 0.07X_2 + 0.13X_3$	-1	+1	-1	-1
$1.65 + 0.07X_2 + 0.13X_3$	+1	-1	-1	-1
$0.65 - 0.07X_2 + 0.13X_3$	-1	-1	+1	+1
$0.61 + 0.07X_2 + 0.1X_3$	-1	+1	-1	+1
$1.89 + 0.07X_2 + 0.13X_3$	+1	-1	-1	+1
$0.19 - 0.07X_2 + 0.13X_3$	-1	+1	+1	-1
$1.39 - 0.07X_2 + 0.13X_3$	+1	-1	+1	-1
$0.51 + 0.07X_2 + 0.13X_3$	+1	+1	-1	-1
$0.43 - 0.07X_2 + 0.13X_3$	-1	+1	+1	+1
$1.63 - 0.07X_2 + 0.13X_3$	+1	-1	+1	+1
$0.75 + 0.07X_2 + 0.13X_3$	+1	+1	-1	+1
$1.17 - 0.07X_2 + 0.13X_3$	+1	+1	+1	-1
$1.41 - 0.07X_2 + 0.13X_3$	+1	+1	+1	+1

* $X_2=IT$, $X_3=MUS$, $X_4=AGE$, $X_5=DB$, $X_7=WG$, $X_9=NT$

TABLE 48

Optimum Configurations for AKP Measure

X_2^*	X_3	X_4	X_5	X_7	X_9	Optimum
1.4142	-1.4142	-1	-1	-1	-1	1.2
1.4142	-1.4142	-1	-1	-1	+1	1.5
-1.4142	-1.4142	-1	-1	+1	-1	0.1
1.4142	-1.4142	-1	+1	-1	-1	0.1
1.4142	-1.4142	+1	-1	-1	-1	1.4
-1.4142	-1.4142	-1	-1	+1	+1	1.4
1.4142	-1.4142	-1	+1	-1	+1	1.3
1.4142	-1.4142	+1	-1	-1	+1	1.6
-1.4142	-1.4142	-1	+1	+1	-1	-0.1
-1.4142	-1.4142	+1	-1	+1	-1	1.1
1.4142	-1.4142	+1	+1	-1	-1	0.2
-1.4142	-1.4142	-1	+1	+1	+1	0.1
-1.4142	-1.4142	+1	-1	+1	+1	1.3
1.4142	-1.4142	+1	+1	-1	+1	0.5
-1.4142	-1.4142	+1	+1	+1	-1	0.9
-1.4142	-1.4142	+1	+1	+1	+1	1.1

* $X_2=IT$, $X_3=MUS$, $X_4=AGE$, $X_5=DB$, $X_7=WG$, $X_9=NT$

search tasks with the largest input timeout, no background music, 8x2 database organization, and a wallet guide, the user added keypresses are minimum.

Optimum Configuration for STA Measure

Table 49 shows the four different functions resulting from each factor level combination of the two discrete variables (i.e., X_4 and X_{10}) shown in Eq. (6). The BC POL subroutine was used to find optimum solutions of these functions since they were not in the form of a first-order equation. The constraint of the continuous variables (i.e., X_1 , X_2 , and X_3) used ranged from -1.4142 to +1.4142 in coded value. The initial starting point used was the center point of the three continuous variables (i.e., $(X_1, X_2, X_3) = (0, 0, 0)$). Table 50 shows the results of these computer runs.

As shown in Table 50, all optimum solutions were obtained at two different design points, i.e., at $(X_1, X_2, X_3) = (-1.4142, -1.4142, -0.9530)$ or $(-1.4142, 1.4142, -0.9530)$. The maximum STA value among the four different functions was 3.74 per trial. The specific factor level combination which generated this maximum value was $(X_1, X_2, X_3, X_4) = (-1.4142, -1.4142, -1.4142, -1) = (120 \text{ wpm}, 2 \text{ secs}, 36.1 \text{ dBA}, \text{College Age})$ or $(X_1, X_2, X_3, X_4) = (-1.4142, 1.4142, -1.4142, -1) = (120 \text{ wpm}, 10 \text{ secs}, 36.1 \text{ dBA}, \text{College Age})$. When a college age user uses the telephone information system with the slowest speech rate, no background music regardless of the amount of input time out, the transcription average score is maximum.

TABLE 49

Four Equations of STA Measure for Different Combinations of Discrete Factors

Equation	X ₄ *	X ₁₀
3.31 -0.15X ₁ +0.02X ₂ -0.40X ₃ -0.21X ₃ ²	-1	-1
3.31 -0.15X ₁ -0.02X ₂ -0.40X ₃ -0.21X ₃ ²	-1	+1
3.01 -0.15X ₁ +0.02X ₂ -0.40X ₃ -0.21X ₃ ²	+1	-1
3.01 -0.15X ₁ -0.02X ₂ -0.40X ₃ -0.21X ₃ ²	+1	+1

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₁₀=SEX

TABLE 50

Optimum Configurations for STA Measure

X_1^*	X_2	X_3	X_4	X_{10}	Optimum
-1.4142	1.4111	-0.9529	-1	-1	3.74
-1.4141	-1.4142	-0.9532	-1	+1	3.74
-1.4139	1.4136	-0.9592	+1	-1	3.44
-1.4141	-1.4142	-0.9532	+1	+1	3.44

* X_1 =SR, X_2 =IT, X_3 =MUS, X_4 =AGE, X_{10} =SEX

Analysis of Optimum Solutions

When a nonconvex function (i.e., neither convex nor concave) is used in the BCPOL subroutine to obtain optimum solutions, the solutions found may not be global optimum solutions within the design region. In other words, a local optimum solution can be found when optimizing a nonconvex function. On the other hand, if a function is convex or concave (or strictly quasi-convex or quasi-concave), then the optimum solution using the BCPOL subroutine is a global optimum within the design region. For more general nonconvex optimization problems, the output from the BCPOL subroutine cannot be treated as a global optimum solution. For such nonconvex functions, the quality of a solution is often dependent upon the choice of an initial starting point, and special partitioning schemes are necessary for searching for a global optimum solution.

The eigen values of the Hessian matrix of a quadratic function can be used to determine the type of a quadratic function (Bazaraa and Shetty, 1979). The entry in row i and column j of the Hessian matrix of a function $f(\mathbf{X})$ is the second partial derivative $\partial^2 f(\mathbf{X}) / \partial x_i \partial x_j$. If all eigen values of the Hessian matrix are non-negative (i.e., the Hessian is positive semidefinite), then the function is convex. If all eigen values are non-positive (i.e., the Hessian is negative semidefinite), then the function is concave. However, if the Hessian matrix is indefinite (i.e., the eigen values are neither all non-negative nor all non-positive), or for that matter, if one is minimizing a concave function (or maximizing a convex function), then the solution found using the search method may not be a global optimum within the design region.

Since the equations for the UAT and AKP measures were affine in terms of X_1 , X_2 , and X_3 , the optimum solutions for UAT and AKP measures were global optima within the design region. However, it was necessary to check for the global optimality of the equations for the TST and STA measures since they were possibly in the form of a second-order equation in terms of X_1 , X_2 , and X_3 . However, the equation for the TST measure was a three-dimensional hyperplane generated by the three independent variables (i.e., X_1 , X_2 , and X_3). Therefore, the optimum solutions were obtained at the boundary of the design region and were moreover global optima. On the other hand, the eigen values for the STA measure were $(\lambda_1, \lambda_2, \lambda_3) = (0, 0, -2.42)$. As a result, the functions for the STA measure were concave, and hence, the maximum values found by the BC POL subroutine were global maxima. Figure 15 shows the solution space for the TST using X_2 (IT) and X_3 (MUS) variables, and Figure 16 shows the solution space for the STA measure using X_1 (SR) and X_3 (MUS) variables.

Summary of Optimum Solutions

Table 51 shows the summary of optimum solutions for the four dependent measures. As shown in Table 51, an optimum configuration of the independent variables can be obtained. First, the amount of input timeout can be set at the lowest value within the design region (i.e., $X_2 = -1.4142 = 2$ seconds). Second, the wallet guide helps the user use the telephone information system effectively. Third, the use of multiple search tasks tends to increase the number of keypresses although they do not affect the optimum performances in terms of the total search time, user added time, and

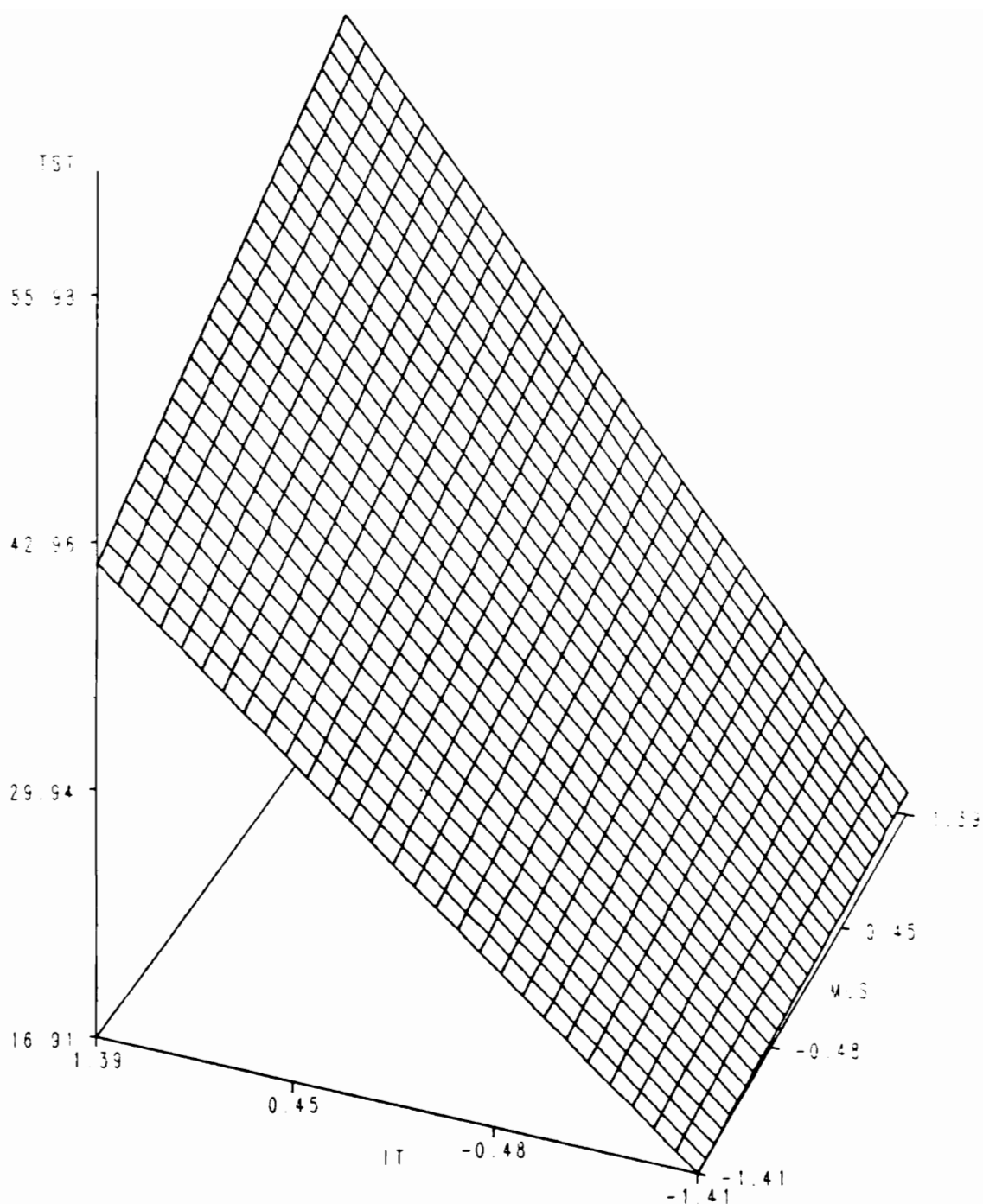


Figure 15. Plot of total search time using input time out and background music level.

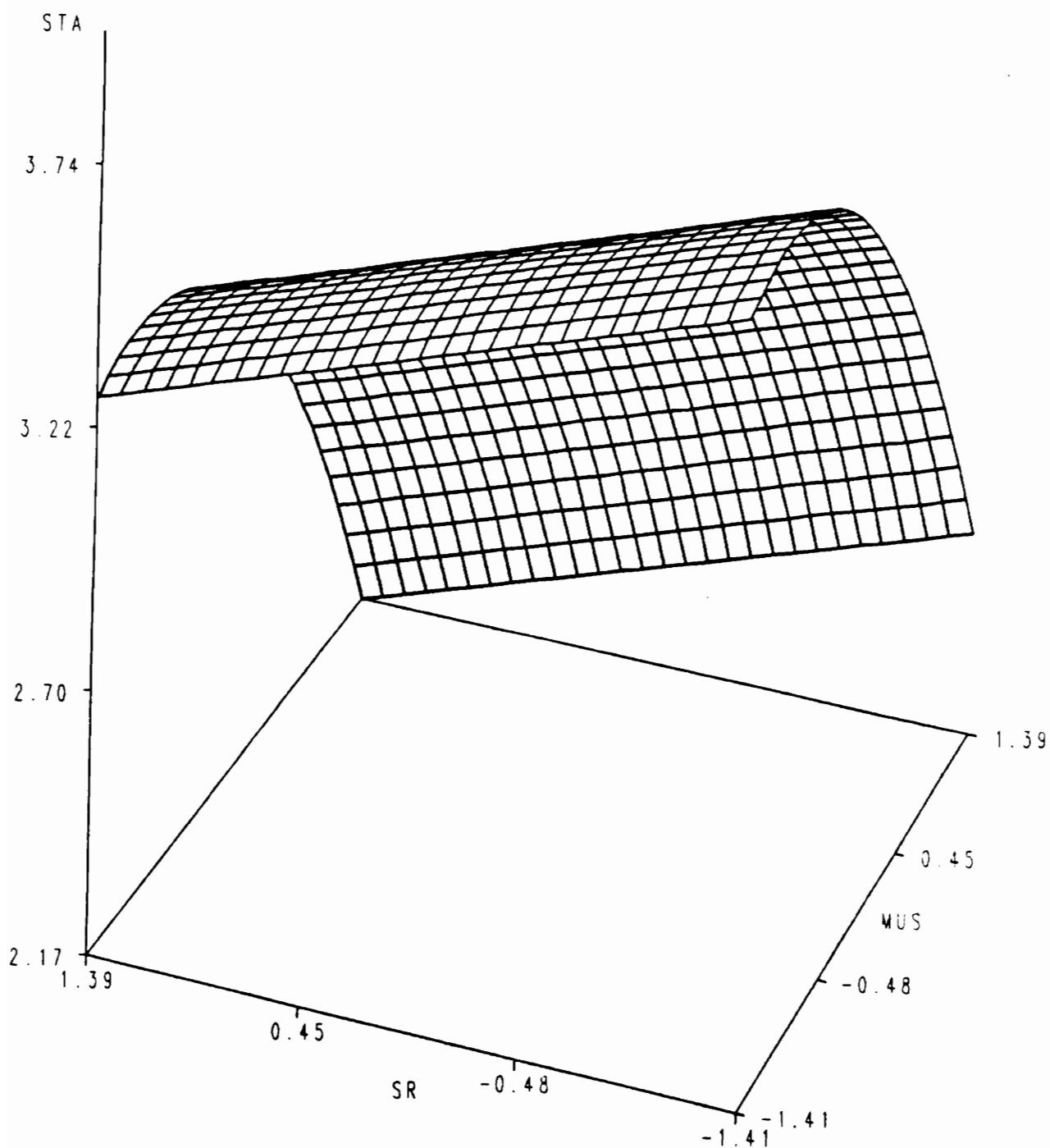


Figure 16. Plot of transcription average scores using speech rate and background music level.

transcription average scores. Finally, the sex of subject, availability of feedback, and type of synthesizer voices do not affect the optimum setting of the telephone information system.

On the other hand, there are several conflicts among the optimum solutions. First, the speech rate of a synthesizer should be set at the fastest value to minimize the total search time. However, the user added time and transcription average scores will be seriously damaged with the fastest speech rate. Second, middle age subjects have smaller user added time than college age subjects although the other dependent measures favor college age subjects. Third, the background music level should be set at the value below -0.95 coded value (i.e., 49.6 dBA) to have optimum performance in terms of transcription average scores although no background music is optimum for the other measures. Finally, the use of 8x2 database structure reduces the number of keypresses to locate a target although the use of 2x6 database structure provides better performance for the other measures.

With these conflicting results, the designer can conduct trade-off analyses to find the optimum setting for the independent variables using the functional relationships. For example, the designer can determine the degree of importance of each dependent measure and determine the optimum setting for the design of a telephone information system.

TABLE 51

Summary of Optimum Solutions

Dependent Measure	X ₁ *	X ₂	X ₃	X ₄	X ₅	X ₇	X ₉
TST	1.41	-1.41	-1.41	-1	-1	1	**
UAT	-1.41	-1.41	-1.41	1	-1	1	**
AKP	**	-1.41	-1.41	-1	1	1	-1
STA	-1.41	-1.41	-0.95	-1	**	**	**

* X₁=SR, X₂=IT, X₃=MUS, X₄=AGE, X₅=DB, X₇=WG, X₈=NT, X₁₀=SEX

** Any value within the design region.

CHAPTER 8

RULES AND CONSTRAINTS OF SEQUENTIAL RESEARCH STRATEGY

This chapter describes general procedures of the sequential research strategy proposed by Williges and Williges (1989). A variety of rules and constraints along with methodological considerations are presented for conducting sequential experiments efficiently. The sequential research strategy primarily deals with the investigation of multiple factors in a systematic and efficient manner. Major procedures consist of: (1) initial selection of potential variables; (2) description of independent variables; and (3) optimization of independent variables as shown in Figure 2. The primary objective of the sequential research strategy is to build second-order empirical models to describe functional relationships among independent variables. This objective can be achieved by combining data sets from several sequential studies which use efficient experimental designs for data collection.

Since this dissertation research deals with the description and optimization of independent variables, major emphases are placed on the rules and constraints in conducting sequential experiments, bridging several data sets resulting from a series of sequential experiments, building integrated empirical models, and optimizing the models. However, a brief summary of

initial selection of potential variables for sequential experiments is also presented based on the paper presented by Merkle, et al. (1988).

Initial Selection

In the first stage of a systematic research strategy, the experimenter needs to identify potential variables which may affect performance of the system. A detailed description of techniques and procedures for selecting initial variables are presented in Chapter 2. Once initial variables are selected, it is necessary to reduce the number of variables since the number of variables is usually so large as to preclude the use of formal experimental designs. Further, there is no need to investigate unimportant variables using extensive data collection based on a formal experimental design. Subjective opinions from both experts and end-users should be substituted for data collection whenever possible since they can be combined meaningfully with the existing scientific literature (Williges and Williges, 1989). When subjective opinions are not available, efficient experimental design procedures are needed to screen the initial variables. The following section describes the use of efficient experimental designs for screening studies.

Screening Study

Screening studies provide an efficient method for screening unimportant variables out of a large number of potential variables. They are a class of fractional factorial designs which allow the experimenter to examine the effects of a large number of variables with a relatively small number of data points (Simon, 1977). These studies assume that all potential variables are not

important in describing the functional relationships between independent variables and their responses. Based on properly designed screening studies, the experimenter is able to determine what variables should be investigated in greater detail. Therefore, the primary purpose of a screening study should be to select important variables but not to obtain complete information on the effects of variables under consideration (Beaudet and Williges, 1988).

A variety of approaches to screening designs have been proposed as a means of screening a large number of factors with a minimum number of data points. One of the most popular approaches is to use a 2^{k-p} fractional factorial design where k is the number of variables of interest and p is the degree of fractionation from the original factorial design. In a 2^{k-p} fractional factorial design, all k factors are studied at only two levels. In choosing a proper 2^{k-p} design, the experimenter must determine design resolution which describes the relationships of confounding structures of the design. In general, a design Resolution R is one in which no p factor effect is confounded with any other effect containing less than $R-p$ factors (Box and Hunter, 1961). For example, in a Resolution III design, no main effects are confounded with any other main effect, but main effects are confounded with two-factor interactions and two-factor interactions with one another. A 2^{3-1} design is of Resolution III. If we assume absence of interactions, k main effects and the overall mean can be orthogonally estimated with only $N = k+1$ runs (provided N is a multiple of 4; otherwise N is rounded upwards to the next multiple of 4, e.g., for $k=2$, $N=4$) (Kleijnen, 1975).

Likewise, a Resolution IV design is one in which no main effects are confounded with any other main effect or two-factor interaction, but where two-

factor interactions are confounded with one another. Therefore, the experimenter can investigate main effects free from being confounded with any two-factor interaction. A 2^{4-1} design is of Resolution IV. Resolution IV designs require $N = 2k$ runs (N being a multiple of 8) to estimate k factors with N runs (Kleijnen, 1975).

These designs, however, are not appropriate to situations with "very" many factors since even a Resolution III design requires at least $k+1$ runs for investigating k factor effects. To provide experimental designs which require fewer runs, supersaturated designs (Booth and Cox, 1962) and group-screening designs (Mauro and Smith, 1982; Watson, 1961) have been developed. Supersaturated designs require fewer experimental runs than the number of factors of interest (i.e., $N < k$). Kleijnen (1975) found that supersaturated designs are more efficient than random designs in which design points are determined by random sampling.

In group-screening designs, factors are divided into groups, each group being considered as a single factor. Then the group-factor may be tested in a 2^{k-p} design as explained in the previous section. If a group factor is found to be insignificant, all original factors within that group are considered insignificant and can be screened out from further investigation. If a group-factor is found to be significant, then one or more original factors within that group are considered significant. Therefore, in the next stage, these original factors should be investigated further. Mauro and Smith (1982) found that these designs performed quite well, both in terms of identifying significant effects and minimizing the number of experimental runs.

Kleijnen (1975) summarized decision criteria in selecting an appropriate design for a screening study. Whenever the experimenter can afford the number of experimental runs required, first-order orthogonal designs (e.g., 2^{k-p} designs) are recommended more than non-orthogonal designs (e.g., supersaturated designs) since they allow the experimenter to investigate main effects unconfounded with each other. However, it may be necessary for the experimenter to evaluate tradeoffs between data collection economy and design properties (such as first-order orthogonality) in choosing an appropriate screening study for a sequential experimental plan.

In summary, the experimenter can select appropriate screening designs depending upon the results of tradeoff analyses between the number of initial variables and resource limitations such as time, budgets, etc. If the number of variables is so large, and hence it is impossible to use a Resolution III design, then the formal procedure for screening initial variables may be: (1) the use of a supersaturated design or a group-screening design; (2) the use of a Resolution III design; and the use of Resolution IV design for selecting final candidate variables for the description stage. However, whenever possible, it is recommended to use a Resolution III design for screening initial variables and then use a Resolution IV design for selecting final candidates. The use of a Resolution IV design is necessary to select final candidates since the primary goal of the sequential research strategy is to build second-order empirical models. In other words, it is necessary to investigate main effects unconfounded with two-factor interactions to select the final candidate variables for the description stage of the sequential research strategy.

Another requirement in conducting screening studies is the choice of a decision criterion (i.e., α level) to accept an effect as being significant. A lax criterion rather than the usual low alpha level (i.e., $\alpha = 0.05$) should be used to prevent variables from being excluded too early in the sequential experiments because no interaction effects among the variables are usually assumed in these screening studies. The experimenter may examine several different criteria to determine which value provides the most meaningful results and satisfies the constraints of resource limitations. The decision criterion level was set at 0.20 in the telephone information study.

Description

Following the screening study, a functional description of the relationships among the selected independent variables needs to be developed to predict human performance in the system. Since the number of variables still may be so large as to preclude the use of complete factorial and multilevel experiments, a series of small experiments should be conducted in an economical and strategic manner (Williges and Williges, 1989). Another reason to conduct a series of experiments rather than a single large experiment is that the set of selected variables resulting from screening studies usually contains two different types of characteristics, i.e., continuous variables and discrete variables. It is difficult to design an experiment, which contains both types of variables, for investigating linear effects and nonlinear effects in an efficient manner. A systematic approach should be made to divide a complete

set of variables into smaller sets on which an efficient estimation of both types of variables is achievable.

Decision criteria for selecting subsets of variables may be established by (1) defining system parameters which will be employed across subsequent studies; (2) determining the number of levels of each variable depending upon the characteristics of information required; (3) characterizing the type of variables (continuous or discrete); and (4) identifying important interactions among variables necessary to be orthogonally investigated.

Before starting formal experiments using subsets of independent variables, the experimenter should establish a variety of experimental procedures or criteria for systematic and consistent investigation of independent variables. These requirements are a critical component in conducting sequential experiments since individual studies should include common characteristics for integrating their results. First, the experimenter should establish a standardized experimental procedure which will be commonly used in subsequent studies during the description stage. If different procedures are used among the subsequent experiments, the data variation resulting from the use of different procedures cannot be estimated in building integrated models. Second, the values of factors, which will be held constant during the experiments, should be clearly recorded for the integration of data sets. Since the integration of several data sets requires systematic knowledge on the levels of each variable (either manipulated or held constant during each individual study), an efficient tool should be developed as a means of administering the results of several studies. Third, dependent measures, which will adequately describe the performance of the system, should include specific measurement

methods. Fourth, there should be common design points across the sequential studies to provide an estimate of comparability of the studies. In other words, the experimenter should carefully design sequential experiments in such a way that there is at least one common treatment condition across the studies. Fifth, system parameters which will be manipulated in all sequential studies should be determined as a means of providing efficiency of data collection while maintaining orthogonality of manipulated factors. The experimenter can select system parameters as those independent variables which will be expected to interact with many other independent variables. Sixth, the experimenter should keep homogeneous samples of subjects across the sequential experiments. In most HCIs, the variability resulting from user differences is usually greater than that of design factors in accounting for performance (Egan, 1988). One way to keep homogeneity of samples is to include important user characteristics in the set of independent variables. These characteristics include users' experience, age, technical aptitudes, and domain specific skills (Egan, 1988). Finally, the experimenter must determine the type of experimental region of interest, e.g., a spherical or cuboidal region generated by continuous variables (Lucas, 1976). If the experimental region is spherical, then all factor levels to examine main effects and two-factor interactions should be set at the factorial portion of a CCD.

Experimental Designs

Once the subsets of independent variables for subsequent small experiments are determined, two-level factorial and fractional factorial designs (i.e., 2^k or 2^{k-p} designs) can be used to evaluate dichotomous variables; whereas CCDs can be used to evaluate continuous variables (Williges and

Williges, 1989). 2^{k-p} fractional factorial designs are an efficient experimental plan when the researcher finds it impossible to collect all the data from a higher-order factorial design due to time and/or budget constraints or, when the researcher can assume higher-order interactions are not present.

A fractional factorial, as the name implies, is a design consisting of a fraction of a complete factorial experiment (Myers, 1976). Although some information will be lost and some effects will be confounded, the fractional replication procedure enables the researcher to determine systematically the confounding structures so that an appropriate subset of treatment combinations can be selected easily. Since many discrete factors in human factors research are dichotomous (e.g., sex, available or not-available feature of a specific factor), 2^{k-p} fractional-factorial designs provide an efficient data collection procedure. For example, the number of design points of a k factorial design is 2^k , whereas a $k-p$ fractional factorial design requires only 2^{k-p} points. When the number of factors is large, the economy of data collection provided by fractional-factorial designs becomes obvious. However, the experimenter should keep Resolution V in constructing a fractional factorial. That is, all main effects and two-factor interactions should be free from being confounded with each other to test their effects orthogonally in each experiment since the primary purpose of the sequential research strategy is to build second-order empirical models.

For the investigation of continuous variables, a second-order polynomial (i.e., main effects, pure quadratics, and two-factor interactions) usually provides an adequate approximation to account for most human performance data (Williges, 1981). To prevent items in the model from being confounded

with each other, an appropriate second-order experimental design should be used.

A CCD provides an experimental plan that allows for efficient estimation of second-order model items. A CCD consists of three portions: (1) 2^k or 2^{k-p} factorial portion; (2) axial portion (represented by α); and (3) center points (Box and Draper, 1987; Myers, 1976). One of the major advantages of the CCD is the economy of data collection. For example, to develop a four-factor, second-order model, a three-level factorial design requires 81 design points, whereas a CCD (one center run without replication) requires only 25 design points. In addition, the CCD allows for sequential experimentation where part of the data are collected and analyzed before proceeding with the additional data collection (Myers, 1976). The factorial portion and the center portion can be used in the preliminary phase to fit a first-order model and to test for the pure quadratic effect. If the pure quadratic effect is obviously significant in the actual response surface, the experimenter can augment the axial points so that the existing results from the original design can be used to fit the second-order model.

In constructing a CCD, several design specifications should be determined. First, the factorial portion should provide at least Resolution V since all main effects and two-factor interactions are specified in the model. Second, the choice of axial points is dependent upon the region of experimental space (cuboidal or spherical) and the design properties such as rotatability and orthogonality. Third, the number of center runs should be determined depending upon the prediction variance property of the design.

Finally, the number of replications should be determined in such a way that an appropriate experimental error can be estimated.

In selecting the levels of factors which will not be manipulated during each individual sequential study, specific levels for the first sequential experiment can be set at the values at which optimum performance is expected. For subsequent studies, factor levels of these constant factors should be consistent with those held constant in the previous sequential studies. For example, suppose a specific factor (say, factor A) was held constant at -1 coded value in the first sequential study. Then, the level of factor A should be set at the same value (i.e., -1 coded value) in the following experiments if these experiments do not manipulate the factor A. In this way, the experimenter can provide common design points across the sequential studies (see Han, et al., (1990) for detailed description of common design points). Also, when a fractional factorial design is used, the experimenter should select carefully a fractional replicate among the alternative fractional replicates. That is, there is always a fractional replicate which contains the common design points if the experimenter designs sequential experiments using the rules described above.

Identifying Uninvestigated Interactions

All main effects and pure quadratic effects are examined during the sequential experiments. However, some two-factor interactions are not examined during the studies since the set of independent variables are divided into several groups. Uninvestigated interactions (two-factor interactions) can be identified when each factor level of the experimental designs is determined. That is, the experimenter can determine which two-factor interactions will not be examined during the sequential experiments before collecting data. These

interactions will be examined by augmenting additional design points with the combined data set. Detailed descriptions for identifying uninvestigated interactions are presented in Chapter 6.

Conducting Sequential Experiments

Using the selected experimental designs and procedures, a series of sequential experiments can be conducted. After collecting data for each sequential experiment, the experimenter can test the effects of manipulated factors for their statistical significance. The usual α level (i.e., $\alpha = 0.05$) can be used in the analysis since the experimental designs have examined the effects of manipulated variables orthogonally. Significant factors (including main effects, two-factor interactions, and pure quadratic effects) should be recorded for the data bridging.

Data Bridging

Comparability of sequential studies. After completing the sequential experiments, the results of each individual studies are combined together. Using the significant effects resulting from the sequential experiments, empirical models can be built. In other words, the empirical models based on the combined data set include only the significant effects from the previous studies. In addition to these significant model items, an artificial factor (called a "block" factor) should be added to the models to estimate comparability of each sequential study. This block factor and the interactions between the block factor and the other existing model items are then examined to determine statistical significance. If the block main effect is significant, the experimenter needs to adjust the results of each individual study. Newman-Keuls tests can be used to determine the adjusting factor of each sequential study. For example, the

mean of sequential study A is twice as large as the mean of sequential study B, then each observation in the sequential study B should be doubled. However, any interaction effect between the block factor and the other existing model items is significant, then the integration of the sequential studies is impossible. In other words, significant interaction effects between the block factor and the other existing model items mean that the results of each sequential study were affected by the other factors which were not included in the set of independent variables. Therefore, the results of each sequential study should be used separately.

Another way of estimating comparability of the sequential studies is to test the performance at the design points commonly included across each individual study. If the performance at the common design points is different across the studies, then the experimenter may have to adjust the results of each sequential study according to the difference, or may not be able to integrate the results of the sequential studies. Further research is needed to determine analysis methods and procedures for examining the common design points.

Selecting additional design points. There are always three different design points for each uninvestigated interaction in the combined data set. Therefore, only one additional design point needs to be added to have four different design points for each uninvestigated interaction. In selecting additional design points, the experimenter needs to identify existing model items which are found to be significant from the previous sequential studies. With these existing model items, several decision rules and criteria are used to select the best design point from the possible candidates for the fourth design

point of each uninvestigated interaction. These decision rules and criteria are summarized as follows (for detailed descriptions of theoretical background and selection procedures, see Appendix J):

- (1) Candidates for each uninvestigated interaction should include the fourth combination of factor level necessary to investigate the interaction effect.
- (2) Determinant values of $X'X$ matrices are computed for the candidates for each uninvestigated interaction.
- (3) The maximum determinant value for each uninvestigated interaction is selected and listed (i.e., If there are n uninvestigated interactions, there are at least n maximum determinant values.). For each uninvestigated interaction, if there are more than one candidate which provide the same maximum values, then candidates which contain the largest number of factor levels of the generating design point are selected.
- (4) The maximum of the maximum determinant values is determined as a final candidate for an additional design point (If more than one candidates exist, an arbitrary candidate among them is selected). The generating design point can be obtained by examining the factor levels of each uninvestigated interaction. That is, the generating design point contains the factor levels which are not included in the combined data set (i.e., the combination of the fourth design point for each uninvestigated interaction).
- (5) This additional design point is added to the combined data set without data collection (i.e., there is no dependent measure observed yet). If this design point consists of all factor levels of the generating design point,

then, rule (6) is applied. If this design point does not contain all factor levels specified in the generating design point, additional design points should be selected until all factor levels of the generating design point are included in the additional design points. At each data point selection, additional design points which have already been selected should be added to the existing model items. The same procedures described in rule (1) through (4) are then applied to selecting additional design points.

- (6) VIF values and eigen value ratios of each uninvestigated interaction are examined. The VIF value of each uninvestigated interaction should be smaller than 10, and the eigen value ratio should be smaller than 1000. If these criteria are not satisfied, additional design points should be selected using the rules (1) through (4) until they are satisfied.

Discussion on constructing $X'X$ matrix. When there are more than one dependent measures, the experimenter should determine what model items should be included in constructing the $X'X$ matrix. It is usually expected that each dependent measure has different model items. As a result, different design points can be selected for each dependent measure. If a large number of different design points is required for each dependent measure, the experimenter may have to sacrifice the optimality of selecting the best design points for each dependent measure. In other words, data collection economy can be significantly reduced with many additional design points. In this case, the experimenter can include all significant existing model items in the $X'X$ matrix across the dependent measures of interest.

Data collection at additional design points. After selecting design points necessary to examine uninvestigated interactions, an additional experiment should be conducted at these design points. With the dependent measures obtained, each uninvestigated interaction is added to the existing model items one at a time. Then partial-F tests are conducted to determine the significance of each uninvestigated interaction effect. Insignificant effects will not be used in developing integrated empirical models. In this way, the experimenter can determine significant interaction effects in an efficient manner.

Empirical Model Building

After examining uninvestigated interactions, the experimenter can add significant interaction effects to the existing model items. However, the integrated empirical models can have serious multicollinearity among the model items by adding more than one additional interactions at the same time. In other words, since each uninvestigated interaction has been tested one at a time under the presence of the existing model items during data bridging, the integrated empirical models should be examined in terms of the degree of multicollinearity. Another reason to examine the integrated models is that the added interaction effects may not be significant any more by adding them together to the existing model items.

To select the best model in terms of variance of coefficients and prediction variances, model selection procedures should be performed. The model selection procedures begin with determining candidate models. Additional rules for determining candidate models can be described as follows (For more detailed descriptions, see Chapter 6):

- (1) If there is no more than one interaction which has been found to be significant during data bridging, additional analyses are not necessary.
- (2) With more than two significant interaction effects, candidate models can be selected by adding all combinations of additional interactions to the existing model items. For example, three candidate models can be selected if there are two additional interactions (say, AB and CD interactions): one with AB interaction added to the existing model items; one with CD interaction added to the existing model items; and finally one with AB and CD interactions added to the existing model items.

With these candidate models, a variety of statistical criteria such as model mean square errors, C_p statistics, PRESS statistics, R-student statistics, average confidence intervals (Myers, 1990) can be used to select the best empirical model in terms of variances of coefficients and prediction variances. After the model selection procedures, the experimenter can obtain the final set of integrated empirical models which describe the functional relationships of performance of the system in the form of second-order empirical models.

Administrative Requirements

Since the sequential research strategy usually requires several studies over a period of time, a procedure for a systematic administration needs to be developed to avoid confusion and lack of consistency in sequential experimentation. In the first stage of research, several decision specifications such as the definition of experimental space, criteria for reducing the number of independent variables, the strategy for screening study, etc. should be clearly established and recorded for subsequent studies. Often, a systematic updating

procedure may be required to record the latest information of changes occurring during the various stages. Following the initial stage, the experimental data points (with constant values of factors) need to be recorded for data integration and optimization procedures. In addition, the results of the sequential studies should be kept in a structured database with specific definitions of dependent measures.

Optimization

To find optimum configurations of the system, a variety of approaches can be used based upon the functional relationships obtained. These include the canonical analysis and ridge analysis (Box, and Draper, 1987; Myers, 1976) and several optimization techniques used in nonlinear programming (Avriel, 1976; Bazaraa and Shetty, 1979). However, since many independent variables in most human factors research are discrete, these techniques cannot be directly applied to optimizing the integrated models which include discrete variables. An alternative approach to optimizing the performance is to use a separate optimization for each combination of discrete variables. In other words, the integrated model can be separately analyzed for each combination of discrete variables and then compared with each other to obtain the treatment level which provides the maximum value of performance.

The canonical analysis and ridge analysis are recommended whenever computer packages are available since they are capable of providing near optimum solutions in addition to optimum solutions. If they are not available, nonlinear programming packages (e.g., IMSL subroutines such as BCPOL,

BCODH, BCOAH, BCLSJ, BCONG, BCLSF, and BCONF) can be used to obtain optimum solutions of the empirical models. These packages require an initial starting point for searching for optimum solutions. When the empirical models are in the form of concave or convex functions, the choice of an initial point is not important since these packages can provide global optimum solutions within the design region of interest. If the empirical models are indefinite (i.e., nonconvex), then the experimenter should examine several initial starting points which are expected to be placed near the optimum solution (For detailed descriptions, see Chapter 7).

With the optimum solutions obtained, the experimenter is able to set optimum configurations of the system and further predict the performance values associated with the optimum solutions. In addition, information about the prediction variances at these optimum solutions are available through the use of statistical packages (e.g., SAS: PROC REG). However, the experimenter should be careful in extrapolating the results beyond the range of the independent variables since these solutions are obtained within the experimental region of interest.

CHAPTER 9

CONCLUSIONS

Integrated Empirical Models of a Telephone Information System

This dissertation research developed integrated empirical models which described the functional relationships among independent variables in the design of a telephone information system. The empirical models were developed using the polynomial regression technique since many of the independent variables affecting human performance at the interface could be considered in a single model. The models used several data sets resulting from a series of related studies based on the sequential research strategy proposed by Williges and Williges (1989). The models for each performance measure (total search time (TST), user added time (UAT), user added keypresses (AKP), and transcription average scores (STA)) are expressed as follows:

$$\begin{aligned} \text{TST} = & 52.81 - 1.20X_1 + 17.27X_2 + 2.77X_3 + 2.15X_4 + 8.27X_5 - 4.47X_7 \\ & + 1.66X_2X_3 + 4.49X_2X_5 - 1.55X_2X_7 + 2.83X_5X_7. \end{aligned}$$

$$\begin{aligned} \text{UAT} = & 9.96 + 0.57X_1 + 1.76X_2 + 3.95X_3 - 3.61X_4 + 0.18X_5 - 4.20X_7 \\ & - 1.72X_2X_5 - 1.38X_2X_7 + 2.29X_5X_7 + 5.45X_4X_5. \end{aligned}$$

$$\text{AKP} = 1.02 + 0.13X_3 + 0.28X_4 - 0.34X_5 - 0.11X_7 + 0.12X_9 - 0.07X_2X_7 \\ + 0.23X_5X_7 + 0.21X_4X_7.$$

$$\text{STA} = 3.16 - 0.15X_1 - 0.40X_3 - 0.15X_4 - 0.02X_2X_{10} - 0.21X_3^2,$$

where X_1 =speech rate, X_2 =input timeout, X_3 =background music level, X_4 =age of subject, X_5 =database organization, X_7 =wallet guide, X_9 =number of target items, and X_{10} =sex of subject. In these equations, an X_i shows the main effect of X_i variable; an X_iX_j shows the linear by linear interaction effect between X_i and X_j ; and an X_i^2 shows the pure quadratic effect of X_i variable. Each coefficient of the model items represents a rate of change when each model item changes by one unit of coded value. The discrete variables (i.e., X_4 through X_{10}) can have either a -1 or a +1 coded value in these equations since they have only two levels (e.g., Available or Not Available); while the continuous variables (i.e., X_1 through X_3) can have any value ranging from -1.4142 to +1.4142 coded values.

From these integrated models, several conclusions can be made. First, the functional relationships show what model items are involved in describing the data variation for each measure. Table 42 lists the model items used for developing the empirical model for each dependent measure. It is interesting to note that three factors (feedback (X_6), type of synthesizer voice (X_8), and sex of subject (X_{10})) resulting from the screening study (Beaudet and Williges, 1988) are not used in developing the empirical models. Seven two-factor interactions out of 45 possible combinations are included in the models. Further, only one

pure quadratic effect (X_3^2) is used. Second, the magnitude of each coefficient represents the relative role of each regressor in describing each dependent measure when each regressor changes by one coded value. For example, input timeout (X_2), interaction between age of subject and database organization (X_4X_5), database organization (X_5), and background music level (X_3) have the greatest effects on describing total search time, user added time, user added keypresses, and transcription average scores respectively. Third, the sign of each coefficient shows the positive (or negative) effect of each regressor. For example, the increase of input timeout (i.e., X_2) increases total search time, while the increase of speech rate (i.e., X_1) decrease total search time.

Using the integrated models, the designer can explore the effects of different combinations of variables and predict human performance. In other words, the designer can determine what variables affect performance of the system, and how much these variables contribute to the performance.

In addition, the integrated models provided estimates of the optimal configuration of the telephone information system (see Table 51). Therefore, the results of the dissertation study can be generalized to estimate the effects of design variations within the specified region of interest (see Table 41), since all significant variables in the design of the telephone information system are considered in the integrated model.

Sequential Research Strategy

The sequential research strategy is expected to have many advantages over a single large experiment when the researcher needs to investigate many independent variables simultaneously. This approach provides a good alternative for investigating multifactors in an efficient manner when higher-order interaction effects are negligible. In addition, it provides a systematic method for investigating important factors and their interactions by introducing the concepts of system parameters and subsets of independent variables. A series of small related experiments also allows the researcher to examine the effects of discrete variables and continuous variables separately using different experimental designs for each type of variables (i.e., 2^{k-p} fractional factorial designs for discrete variables and CCDs for continuous variables).

The use of several types of experimental designs provides a more economical data collection procedure than a single large experiment. For example, to build a complete second-order empirical model for the ten independent variables, a single factorial design ($2^6 \times 3^4$; six dichotomous variables and four continuous variables) would require 5184 design points (without replications); whereas the sequential research strategy required 152 design points (50 (from the Experiment 1) + $3 \times 2^{5-1}$ (from the Experiment 2) + $3 \times 2^{5-1}$ (from the Experiment 3) + 6 (from the Experiment 4)) even with replications. The only advantage of the single factorial design is that the effects of higher-order interactions can be estimated orthogonally.

Although the resulting data sets from the series of studies are not orthogonal when integrated together, this strategy can provide an efficient way

to interpret the integrated data set. That is, since the researcher uses resolution V designs during the description stage, the effects of important factors can be estimated orthogonally. In fact, the results of a simulation study show that the integrated data set is near-orthogonal because the uninvestigated interactions are partially confounded with other factors.

The lessons learned from this study provided guidelines and procedures in integrating several data sets based on sequential experimentation. The integrated data set has important uses in human factors research because it can provide a generalizable result in the design of a specific system. In addition, this study provided guidelines for efficient procedures for planning and conducting sequential experiments in human factors research. These procedures can be used as an efficient strategy in investigating a large number of independent variables in a systematic manner. Decision rules and criteria in constructing a research plan were included in the guidelines as a means of providing an efficient tool for obtaining generalizable results. Figure 17 shows a brief summary of procedures and methodological considerations in conducting sequential experiments.

Future Research

The following is a brief summary of future research for improving the sequential research strategy.

- Several data sets result from the sequential research strategy. These include intermediate results from each study, data from reduction

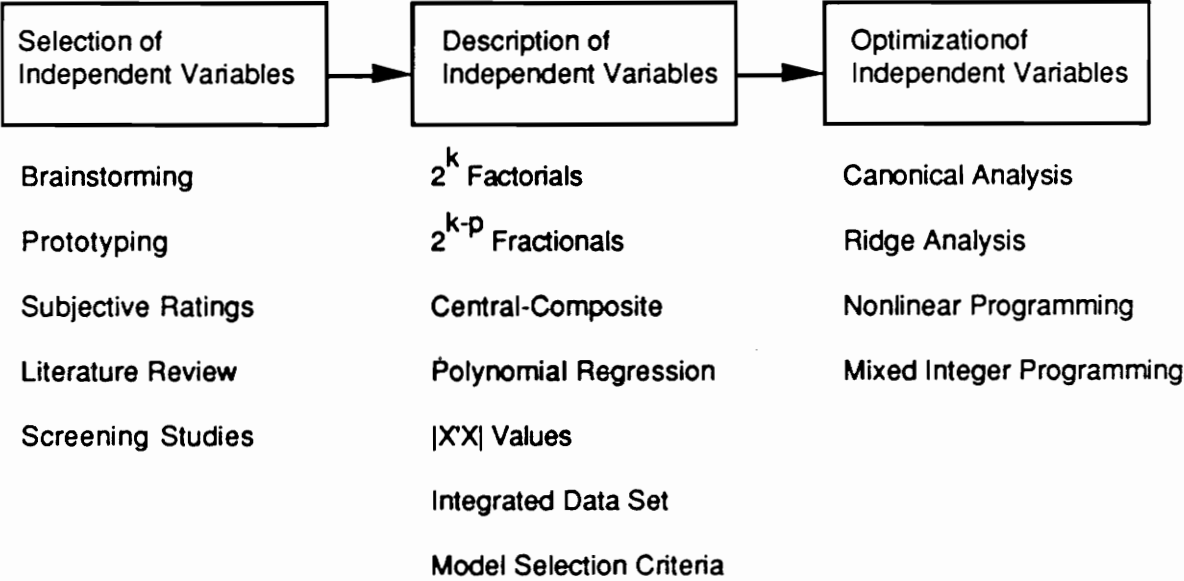


Figure 17. Tools for conducting sequential experiments (Adapted from Williges and Williges, 1989).

procedures, and the final results for analysis. If these data sets are not organized in a systematic manner, it may be difficult to integrate them into a combined one. Tools for administering research data sets in an efficient and accurate manner need to be developed.

- Although several suggestions were made for dividing the set of independent variables for the description stage, more research is needed to simulate what methods provide systematic and efficient procedures to reduce the number of data points required to conduct sequential experiments; while satisfying desirable statistical criteria (e.g., orthogonality, prediction variances, etc.).
- Before starting sequential experiments, the experimenter should consider another alternative for examining many independent variables. Resolution V fractional factorial designs can be used to examine the effects of independent variables. Curvalinear linear effects generated by continuous variables can then be examined by adding center points of the continuous variables. This approach may be useful when there is no pure quadratic effects. More research is necessary to determine the levels of discrete factors at the center points, tradeoffs between this approach and the method suggested above in terms of data collection economy, statistical properties such as orthogonality.
- Since the sequential research strategy requires manipulation of many variables in a series of related studies, the effects of the manipulated variables in a study should be reflected on the subsequent studies depending upon the significance of the effects. The experimenter must set

clear stopping points for terminating sequential studies when the results are not fruitful.

- Testing methods for examining comparability of sequential studies need to be developed. These include tools for analyzing the data resulting from common design points and methods for adjusting the results when there are significant differences among the studies unexplained by the manipulated variables.
- Although this research suggests a way to determine the levels of factors which will be held constant during a sequential study, there may be other alternatives such as using different levels of constant factors during each study. These alternatives should be examined in terms of obtaining better statistical properties (e.g., orthogonality) and providing common design points which may not be included in formal experimental designs for sequential studies.
- When selecting additional design points to examine uninvestigated interactions, efficient ways to reduce multicollinearity should be developed. Alternatives include replications at the same design points selected or adding new design points to the integrated data set. These two methods should be examined to determine which method provides a more efficient way to reduce multicollinearity.
- The proposed separate optimization procedure is inefficient when the number of discrete variables is large. The use of techniques such as mixed integer programming (Lazimy, 1985) to obtain optimum solutions which include both continuous and discrete variables should be examined.

- If a nonlinear programming technique is used to obtain optimum solutions, specific rules need to be developed for selecting initial starting points.

REFERENCES

- Avriel, M. (1976). Nonlinear programming: Analysis and methods. Englewood Cliffs, NJ: Prentice-Hall.
- Bazaraa, M. S. and Shetty, C. M. (1979). Nonlinear programming: Theory and algorithms. New York, NY: John Wiley.
- Beaudet, B. B. (1988). The effects of 16 variables on a telephone information system which uses synthetic speech. Unpublished Master's Thesis. Virginia Polytechnic Institute and State University. Blacksburg, VA.
- Beaudet, D. B. and Williges, R. C. (1988). The role of screening studies in sequential research designs. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 1174-1178). Santa Monica, CA: Human Factors Society.
- Bennett, J. L. (1984). Managing to meet usability requirements. In J. Bennett, D. Case, J. Sandlin, and M. Smith (Eds.), In visual display terminals. Englewood Cliffs, NJ: Prentice-Hall.
- Berney, C. L. (1984). Text-to-speech in the electronic mail environment. In Proceedings of Speech Tech '84 (pp. 27-31). New York, NY: Media Dimensions.
- Booth, K. H. V. and Cox, D. R. (1962). Some systematic supersaturated designs. Technometrics, 4, 489-495.
- Box, G. E. P. and Draper, N. R. (1987). Empirical model-building and response surfaces. New York, NY: John-Wiley.

- Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} fractional factorial designs. Technometrics, 3(3), 311-351.
- Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). Statistics for experimenters. New York, NY: Wiley.
- Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions. Journal of the Royal Statistical Society. Series B (Methodological), 1-45.
- Carroll, J. M., Landauer, T. K., John, B. E., Whiteside, J., and Wolf, C. G. (1989). The role of laboratory experiment in HCI: Help, hindrance, or ho-hum? (PANEL). In Proceedings of CHI'89 Human Factors in Computing Systems (pp. 265-268). New York, NY: ACM.
- Chapanis, A. (1988). Some generalizations about generalization. Human Factors, 30(3), 253-267.
- Chapanis, A. and Van Cott, H. P. (1972). Human engineering tests and evaluation. In H. P. Van Cott and R. G. Kinkade (Eds.) Human engineering guide to equipment design. New York, NY: Wiley .
- Deffner, G. and Melder, K. (1990). User acceptance and preference for advanced voice services features and dialogue styles. In Proceedings of the Human Factors Society 34th Annual Meeting (pp. 194-197). Santa Monica, CA: Human Factors Society.
- Detweiler, M. C., Schumacher, R. M. Jr., and Gattuso, N. L., Jr. (1990). Alphabetic input on a telephone keypad. In Proceedings of the Human Factors Society 34th Annual Meeting (pp. 212-216). Santa Monica, CA: Human Factors Society.

- Diamond, W. J. (1981). Practical experiment designs for engineers and scientists. Belmont, CA: Lifetime Learning.
- Dykstra, O., Jr. (1971). The augmentation of experimental data to maximize $|X'X|$. Technometrics, 13(3), 682-688.
- Egan, D. E. (1988). Individual differences in human-computer interaction. In M. Helander (Eds.) Handbook of human-computer interaction (pp. 543-568). Elsevier, North-Holland.
- Evans, J. W. (1979). Computer augmentation of experimental designs to maximize $|X'X|$. Technometrics, 21(3), 321-330.
- Gill, P. E., Murray W., and Wright M. H. (1981). Practical optimization. New York, NY: Academic Press.
- Gould, J. D. and Boies, S. J. (1983). Human factors challenges in creating a principal support office system - The speech filing system approach. ACM Transactions on Office information Systems, 30, 273-298.
- Gould, J. D. and Boies, S. J. (1984). Speech filing - An office system for principles. IBM Systems Journal, 23(1), 65-81.
- Gould, J. D. and Lewis, C. (1983). Designing for usability: Key principles and what designers think. In Proceedings of CHI'83 Human Factors in Computing Systems (pp. 50-53). New York, NY: ACM.
- Gould, J. D. and Lewis, C. (1985). Designing for usability: Key principles and what designers think. Communications of ACM, 28(3), 300-311.
- Greene, B. G., Logan, J. S., and Pisoni, D. B. (1986). Perception of synthetic speech produced automatically by rule: Intelligibility of eight text-to-speech systems. Behavior Research Methods, Instruments, and Computers, 18, 100-107.

- Greene, B. G., Manous, L. M., and Pisoni, D. B. (1984). Perceptual evaluation of 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. In Proceedings of CHI'89 Human Factors in Computing Systems (pp. 347-352). New York, NY: ACM.
- Han, S. H., Williges, R. C., and Williges, B. H. (1990). Empirical models based on data from several experiments. In Proceedings of the Human Factors 34th Annual Meeting (pp. 1224-1228). Santa Monica, CA: Human Factors Society.
- Hebble, T. L. and Mitchell, T. J. (1972). "Repairing" response surface designs. Technometrics, 14(3), 767-779.
- Herlong, D. W. and Williges, B. H. (1988). Designing speech displays for telephone information systems. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 215-218). Santa Monica, CA: Human Factors Society.
- 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.
- Kleijnen, J. P. C. (1975). Screening designs for poly-factor experimentation. Technometrics, 17(4), 487-493.

- Landauer, T. K. (1988). Research method in human-computer interaction. In M. Helander (Eds.), Handbook of human-computer interaction (pp. 905-928). Elsevier, North-Holland.
- Lauretta, D. J., Rodman, R. D., and Antin, J. F. (1990). The effects of familiarization on the comprehension of synthetic speech in telephone communication. In Proceedings of Human Factors Society 34th Annual Meeting (pp. 189-193). Santa Monica, CA: Human Factors Society.
- Lazimy, R. (1985). Improved algorithm for mixed-integer quadratic programs and a computational study. Mathematical Programming, 32, 100-113.
- Lucas, J. M. (1976). Which response surface design is best? Technometrics, 18, 411-417.
- Marics, M. A. (1990). How do you enter "D'ANZI-Quist" using a telephone keypad? In Proceedings of the Human Factors Society 34th Annual Meeting (pp. 208-211). Santa Monica, CA: Human Factors Society.
- Martin, M. M., Williges, B. H., and Williges, R. C. (1990). Improving the design of telephone-based information systems. In Proceedings of the Human Factors 34th Annual Meeting (pp. 198-202). Santa Monica, CA: Human Factors Society.
- Mauro, C. A. and Smith, D. E. (1982). The performance of two-stage group screening in factor screening experiments. Technometrics, 24(4), 325-330.
- Merkle, P. J., Jr. (1988). Using subjective ratings to select independent variables in the design of telephone inquiry systems. Unpublished

Master's Thesis. Virginia Polytechnic Institute and State University.
Blacksburg, VA.

Merkle, P. J., Jr., Beaudet, D. B., Williges, R. C., Herlong, D. W., and Williges, B. H. (1988). A systematic method for the selection of independent variables in the investigation of complex systems. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 1179-1182). Santa Monica, CA: Human Factors Society.

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

Mills, C. B. (1986). Usability testing in the real world. (PANEL). In Proceedings of CHI'86 Human Factors in Computing Systems (pp. 212-215). New York, NY: ACM.

Mitchell, T. J. (1974a). Computer construction of "D-optimal" first-order designs. Technometrics, 16(2), 211-230.

Mitchell, T. J. (1974b). An algorithm for the construction of "D-optimal" experimental designs. Technometrics, 16(2), 203-210.

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

Myers, R. H. (1976). Response surface methodology. Boston, MA: Allyn and Bacon.

Myers, R. H. (1990). Classical and modern regression with applications (2nd Eds). Boston, MA: PWS-Kent.

- Podgorny, P. (1985). Telephone as computer terminal. In Official Proceedings of Speech Tech '85 (pp. 103-109). New York, NY: Diamond Dimensions.
- Ray, W. J. and Ravizza, R. (1986). Methods toward a science of behavior and experience (2nd Eds.). Belmont, CA: Wadsworth Publishing Company.
- Sanders, M. S. and McCormick, E. J. (1987). Human factors in engineering design. New York, NY: McGraw-Hill.
- SAS Institute Inc. (1986). Statistical analysis system, release 5.16. Cary, NC: SAS Institute Inc.
- Schmandt, C. W. (1985). Voice access to an electronic mail system. In Official Proceedings of Speech Tech '85 (pp. 89-91). New York, NY: Media Publications.
- Shackel, B. (1984). The concept of usability. In J. Bennett, D. Case, J. Sandlin, and M. Smith (Eds.), In visual display terminals. Englewood Cliff, NJ: Prentice-Hall.
- Simon, C. W. (1973). Economical multifactor designs for human factors engineering experiments (Tech Report. P73-326A). Culver City, CA: Huges Aircraft Company.
- Simon, C. W. (1977). Design, analysis, and interpretation of screening studies for human factors engineering (Tech. Report CWS-03-77). Culver City, CA: Huges Aircraft Company.
- Simon, C. W. (1987). Will egg-sucking ever become a science? In Human Factors Society Bulletin (pp. 1-4). 30(6), Santa Monica, CA: Human Factors Society.

- Simpson, C. A., McCauley, M. E., Roland, E. F., Ruth, J. C., and Williges, B. H. (1985). System design for speech recognition and generation. Human Factors, 27, 115-141.
- Smith, L. L. (1987). Whyfore human factors? In Human Factors Society Bulletin (pp. 6-7). 30(4), Santa Monica, CA: Human Factors Society.
- Sorkin, R. D. (1987). Design of auditory and tactile displays. In G. Salvendy (Eds.), Handbook of human factors (pp. 549-576). New York, NY: Wiley-Interscience.
- Spine, T. M., Williges, B. H., and Maynard, J. C. (1984). An economical approach to modeling speech recognition accuracy. International Journal of Man-Machine Studies, 21, 191-202.
- Streeter, L. A. (1988). Applying speech synthesis to user interfaces. In M. Helander (Eds.), Handbook of human-computer interaction (pp. 321-343). Elsevier, North-Holland.
- Thimbleby, H. (1990). User interface design. New York, NY: ACM.
- Thomas, J. C., Rosson, M. B., and Chodorow, M. (1984). Human factors and synthetic speech. In Proceedings of the Human Factors Society 28th Annual Meeting (pp. 763-767). Santa Monica, CA: Human Factors Society.
- Watson, G. S. (1961). A study of the group-screening method. Technometrics, 3(3), 371-388.
- Whiteside, J., Bennett, J., and Holtzblett, K. (1988). Usability engineering: Our experience and evolution. In M. Helander (Eds.) Handbook of human-computer interaction (pp. 791-817). Elsevier, North-Holland.

- Williams, K. (1978). An introduction to the assessment and measurement of sound. In D. M. Lipscomb and A. C. Taylor (Eds.), Noise control: Handbook of principles and practices (pp. 33-60). New York, NY: Van Nostrand Reinhold.
- Williges, B. H., Schurick, J. M., Spine, T. M., and Hakkinen, M. T. (1986). Using speech in the human computer interface. In R. Ehrich and R. Williges (Eds.), Advanced in human factors/ergonomics (vol 2.): Human-computer dialogue design (pp. 241-271). New York, NY: Elsevier.
- Williges, B. H. and Williges, R. C. (1984). Dialogue design considerations for interactive computer systems. In F. Muckler (Eds.), Human factors review: 1984 (pp. 167-208). Santa Monica, CA: Human Factors Society.
- Williges, R. C. (1981). Development and use of research methodologies for complex system/simulation experimentation. In M. Morral and K. Kraiss (Eds.), Manned system design (pp. 59-87). New York, NY: Plenum.
- Williges, R. C. (1987). The use of models in human-computer interface design. Ergonomics, 30(3), 491-502.
- Williges, R. C. and Ehrich, R. W. (1986). Multidisciplinary research in human-computer dialogue design. In R. Ehrich and R. Williges (Eds.), Advanced in human factors/ergonomics (vol 2.): Human-computer dialogue design (pp. 1-9). New York, NY: Elsevier.

- Williges, R. C. and Mills, R. G. (1982). Catalog of methodological considerations for systems experimentation (Tech. Report). Wright-Patterson Air Force Base, OH: Aeronautical Systems Division.
- Williges, R. C. and Williges, B. H. (1982). Modeling the human operator in computer based data entry. Human Factors, 24(3), 285-299.
- Williges, R. C. and Williges, B. H. (1989). Integrated research paradigm for complex experimentation. In Proceedings of the Human Factors Society 33th Annual Meeting, (pp. 606-610). Santa Monica, CA: Human Factors Society.
- Williges, R. C., Williges, B. H., and Elkerton, J. (1987). Software interface design. In G. Salvendy (Eds.), Handbook of human factors (pp. 1416-1449). New York, NY: Wiley.
- Witten, I. H. and Madams, 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 human performance trade-offs of a telephone-based information system. Unpublished Master's Thesis. Virginia Polytechnic Institute and State University. Blacksburg, VA.

Appendix A
Constant Factor Levels

Constant Factor Levels
(From Merkle, 1988)

Variable	Level
Mean Pitch	Paul (120 Hz)
Range of Pitch	Paul (100 %)
Assertiveness	Paul (100 %)
Richness	Paul (20 %)
Head size - Resonance	Paul (100 %)
Gain - Fricatives	Paul (73 dB)
Gain - Aspiratives	Paul (70 dB)
Gain - Voicing	Paul (71 dB)
Gain - Nasal	Paul (69 dB)
Sex of Speaker	Paul (male)
Transaction Summary	none
Hardcopy Summary	none
Amplitude Control on Telephone	none
Sex of Listener (User)	male/female
Overall Speech Rate	Paul
Smoothness	Paul
Breathiness	Paul
Familiarity of Words	familiar
Active vs. Passive Sentences	active
Simple vs. Complex Sentences	mixed (report)
Length of Messages	1 sentence
Order of Information Message	end
Double Keying	none
Number of Commands	3 to 12, 4 recommended
Error Detection	none
Undoing Actions	none
Embedded Training	none
Repeat Display	keywords(0), messages(0)
Hearing Impairment	none
Native vs. Foreign Listener	both
Competing Speech	none
Noise	ambient
Length of Sentences	noun verb-phrase object
Order of Information	end
Keypress Echoing	none

**Constant Factor Levels
(Continued)**

Variable	Level
Size of vocabulary	report
Length of Words	report
Use of Jargon	none
System Intelligence	none
Methods of Transversal	keyword
Length of Commands	one
Command Abbreviations	none
Command Synonyms	none
Error Handling	none
Change Wording or Phrasing	none
Change Rate of Speech Display	none
Initiation of Help	user
Selection of Help	user
Content of Help	annotation
Access of Help	direct
Organization of Help	linear
Human Assistance	none
Frequency Control	none
Repeat Word or Sentence	none
Suppressing Menu Prompts	none
User Defined Markers	none
Experience with Computers	report
Experience with Speech Synthesis	report
Experience with Other Speech	report
Type of Data	verbal and numerical
Number of Searches and Sessions	two
Logical Operators	none
If/Then Target Selection	none
General vs. Specific Search	specific
Type of Competing Tasks	none

Appendix B
Experimental Data Points and Dependent Measures

Raw Data Set

OBS#	SR	IT	MUS	AGE	DB	FB	WG	TV	NT	SEX	TST	UAT	AKP	STA	EXP
1		1	1	-1	1	-1	-1	1	-1	1	107.36	36.984	1.313	2.9375	1
2		1	-1	-1	1	-1	-1	1	-1	1	36.774	8.386	0.063	3.25	1
3		1	1	-1	1	-1	-1	1	-1	1	79.711	9.319	0.125	3.25	1
4		1	-1	-1	1	-1	-1	1	-1	-1	32.694	4.293	0.25	3.3125	1
5		-1	1	-1	1	-1	-1	1	-1	-1	93.26	18.617	1.125	2.8125	1
6		0	0	-1	1	-1	-1	1	-1	-1	54.566	3.448	0.188	3.3125	1
7		-1	1	-1	1	-1	-1	1	-1	1	76.144	1.522	0	3.437	1
8		-1	-1	1	1	-1	-1	1	-1	-1	34.862	2.245	0.188	2.625	1
9		-1	-1	-1	1	-1	-1	1	-1	-1	38.733	6.108	0.25	3.375	1
10		1	1	-1	1	-1	-1	1	-1	-1	78.043	7.667	0.063	2.375	1
11		1	1	-1	-1	1	-1	1	-1	1	72.19	1.799	0.063	3.5	1
12		1	-1	1	-1	1	-1	1	-1	1	36.573	8.184	0.438	1.875	1
13		1	-1	-1	1	-1	-1	1	-1	1	36.693	8.292	0.375	3.1875	1
14		0	0	-1	1	-1	-1	1	-1	1	60.097	8.979	0.188	3.1875	1
15		-1	1	-1	1	-1	-1	1	-1	1	77.903	3.26	0.313	2.5	1
16		-1	1	-1	-1	1	-1	1	-1	-1	76.042	1.42	0	3.75	1
17		-1	-1	1	-1	1	-1	1	-1	-1	40.902	8.284	0	2.25	1
18		-1	-1	-1	1	-1	-1	1	-1	1	36.648	4.024	0	3.625	1
19		0	-1.414	0	1	1	-1	1	-1	-1	34.047	10.962	0.626	3.4375	1
20		1.4142	0	0	1	1	-1	1	-1	-1	71.242	22.103	0.5	2.375	1
21		-1.414	0	0	1	1	-1	1	-1	-1	55.703	1.149	0	3.8125	1
22		0	0	0	1	1	-1	1	-1	1	99.299	48.218	1.001	1.5	1
23		0	1.4142	0	1	1	-1	1	-1	1	80.073	0.992	0.063	3.3125	1
24		0	0	1.4142	1	1	-1	1	-1	-1	63.924	12.838	0.438	1.5	1
25		0	0	-1.414	1	1	-1	1	-1	1	52.922	1.819	0	3.1875	1
26		0	0	0	1	1	-1	1	-1	1	58.898	7.834	0.188	3.6875	1
27		0	0	0	1	1	-1	1	-1	1	52.15	1.086	0	3.25	1

Raw Data Set (Continued)

OBS#	SR	IT	MUS	AGE	DB	FB	WG	TV	NT	SEX	TST	UAT	AKP	STA	EXP #
28		0	0	0	1	1	-1	1	-1	1	57.679	6.598	0.063	3.375	1
29		1	1	-1	1	1	-1	1	-1	-1	75.101	4.717	0.188	3.25	1
30		0	-1.414	0	1	1	-1	1	-1	1	42.732	19.647	0.438	3.4375	1
31	1.4142	0	0	0	1	1	-1	1	-1	-1	57.999	8.86	0.563	2.1875	1
32		1	-1	-1	1	1	-1	1	-1	-1	53.71	25.315	1.563	2.8125	1
33	-1.414	0	0	0	1	1	-1	1	-1	-1	62.268	7.714	0.125	3.625	1
34		0	1.4142	0	1	1	-1	1	-1	1	83.246	4.165	0.125	3.25	1
35		0	0	-1.414	1	1	-1	1	-1	1	59.483	8.38	0.501	3.6875	1
36		1	-1	1	1	1	-1	1	-1	-1	31.98	3.583	0.063	2.3125	1
37		-1	-1	-1	1	1	-1	1	-1	-1	46.981	14.351	0.625	2.75	1
38		0	0	1.4142	1	1	-1	1	-1	-1	65.099	14.013	0.375	1.8125	1
39		-1	1	-1	1	1	-1	1	-1	1	81.811	7.179	0.125	2.4375	1
40		1	1	-1	1	1	-1	1	-1	-1	86.015	15.63	0.313	2.9375	1
41		1	1	1	1	1	-1	1	-1	1	161.28	90.892	3	2.0625	1
42		-1	-1	1	1	1	-1	1	-1	-1	33.938	1.334	0.063	3.375	1
43		-1	-1	1	1	1	-1	1	-1	-1	39.423	6.819	1.438	3	1
44		1	-1	1	1	1	-1	1	-1	-1	62.261	33.865	2.188	1.75	1
45		-1	-1	-1	1	1	-1	1	-1	1	33.99	1.359	0	3.125	1
46		-1	1	1	1	1	-1	1	-1	-1	91.854	17.228	0.625	3.5	1
47		-1	1	1	1	1	-1	1	-1	-1	92.019	17.394	0.688	2.375	1
48		1	1	1	1	1	-1	1	-1	-1	89.463	19.072	0.813	2.4375	1
49		1	-1	-1	1	1	-1	1	-1	-1	43.467	15.072	0.438	2.875	1
50		-1	1	-1	1	1	-1	1	-1	-1	76.004	1.372	0	2.9375	1
51	-1.414	1.4142	-1.414	-1	-1	-1	1	1	-1	-1	83.197	41.522	3.938	3.625	2
52	-1.414	-1.414	-1.414	-1	-1	-1	1	1	-1	1	30.978	13.451	1.063	3.5625	2
53	-1.414	-1.414	-1.414	-1	-1	1	1	1	-1	1	27.589	0.377	0	3.6875	2
54	-1.414	-1.414	-1.414	-1	-1	1	1	1	-1	1	35.159	7.947	0.438	3.75	2

Raw Data Set (Continued)

OBS#	SR	IT	MUS	AGE	DB	FB	WG	TV	NT	SEX	TST	UAT	AKP	STA	EXP #
55	1.4142	1.4142	-1.414	-1	-1	1	1	1	-1	-1	39.264	3.173	0.125		3
56	-1.414	1.4142	-1.414	-1	1	1	1	1	-1	-1	84.357	1.023	0	3.5625	2
57	-1.414	-1.414	-1.414	-1	1	1	-1	1	-1	1	29.689	2.477	0.125	3.875	2
58	-1.414	1.4142	-1.414	-1	-1	1	-1	1	-1	1	60.702	19.028	0.5	3.1875	2
59	1.4142	-1.414	-1.414	-1	1	-1	-1	1	-1	1	23.839	2.001	0.125	3.125	2
60	1.4142	-1.414	-1.414	-1	1	1	1	1	-1	-1	27.068	5.296	0.938	3.0625	2
61	-1.414	-1.414	-1.414	-1	-1	-1	-1	1	-1	-1	23.142	5.616	0.375	3.6875	2
62	-1.414	-1.414	-1.414	-1	-1	-1	-1	1	-1	1	36.75	19.223	1.125	3.6875	2
63	1.4142	-1.414	-1.414	-1	1	1	1	1	-1	1	22.684	0.913	0	3.25	2
64	-1.414	1.4142	-1.414	-1	1	-1	-1	1	-1	-1	99.173	15.893	0.688	3.625	2
65	-1.414	1.4142	-1.414	-1	-1	-1	1	1	-1	-1	46.472	4.909	0.063	3.625	2
66	1.4142	1.4142	-1.414	-1	-1	-1	-1	1	-1	-1	60.091	24.049	1.063	3.1875	2
67	-1.414	1.4142	-1.414	-1	-1	-1	1	1	-1	-1	46.642	5.078	0.375	3.625	2
68	-1.414	-1.414	-1.414	-1	1	-1	1	1	-1	-1	30.928	3.626	0	3.25	2
69	1.4142	-1.414	-1.414	-1	1	1	1	1	-1	1	24.689	2.918	0.188	3.8125	2
70	-1.414	-1.414	-1.414	-1	1	-1	1	1	-1	-1	32.944	5.642	0.125	3.625	2
71	-1.414	1.4142	-1.414	-1	1	-1	-1	1	-1	-1	84.019	0.739	0	3.6875	2
72	-1.414	-1.414	-1.414	-1	1	-1	1	1	-1	1	27.585	0.283	0	3.9375	2
73	1.4142	-1.414	-1.414	-1	-1	-1	1	1	-1	-1	17.721	5.705	0.876	3.4375	2
74	-1.414	-1.414	-1.414	-1	-1	1	1	1	-1	1	20.082	2.424	0	3.8125	2
75	1.4142	1.4142	-1.414	-1	-1	-1	-1	1	-1	-1	54.539	18.497	2.625	3.4375	2
76	1.4142	-1.414	-1.414	-1	-1	1	-1	1	-1	1	30.291	18.194	0.438	3.125	2
77	1.4142	1.4142	-1.414	-1	-1	1	1	1	-1	1	41.367	5.277	0	3.5	2
78	1.4142	1.4142	-1.414	-1	1	-1	1	1	-1	1	79.796	1.998	0	3.0625	2
79	1.4142	1.4142	-1.414	-1	1	1	-1	1	-1	-1	85.057	7.28	0.126	3.4375	2
80	-1.414	1.4142	-1.414	-1	1	1	1	1	-1	-1	83.616	0.282	0	3.4375	2
81	-1.414	1.4142	-1.414	-1	-1	1	-1	1	-1	-1	88.944	47.27	2	3.3125	2

Raw Data Set (Continued)

OBS#	SR	IT	MUS	AGE	DB	FB	WG	TV	INT	SEX	TST	UAT	AKP	STA	EXP#
82	1.4142	-1.414	-1.414	-1	1	-1	-1	1	-1	1	24.078	2.24	0	2.5625	2
83	1.4142	-1.414	-1.414	-1	1	-1	-1	1	-1	1	24.611	2.773	0.251	3.6875	2
84	1.4142	1.4142	-1.414	-1	-1	1	1	1	-1	-1	38.959	2.869	0	3.25	2
85	-1.414	1.4142	-1.414	-1	1	1	1	1	-1	-1	83.894	0.56	0	3.8125	2
86	-1.414	1.4142	-1.414	-1	1	-1	-1	1	-1	1	91.331	8.051	0.438	3.625	2
87	-1.414	-1.414	-1.414	-1	-1	1	1	1	-1	1	19.251	1.593	0.063	3.875	2
88	1.4142	1.4142	-1.414	-1	1	1	-1	1	-1	1	80.424	2.647	0.063	3.4375	2
89	-1.414	-1.414	-1.414	-1	-1	1	1	1	-1	1	27.819	10.161	0.563	3.3125	2
90	-1.414	1.4142	-1.414	-1	-1	1	1	1	-1	1	44.457	2.893	0.125	3.6875	2
91	1.4142	1.4142	-1.414	-1	1	-1	1	1	-1	1	79.916	2.118	0	3.3125	2
92	1.4142	-1.414	-1.414	-1	-1	-1	1	1	-1	-1	17.466	5.451	0.063	3.125	2
93	1.4142	1.4142	-1.414	-1	1	1	-1	1	-1	-1	78.797	1.02	0.063	3.1875	2
94	1.4142	1.4142	-1.414	-1	-1	-1	-1	1	-1	-1	49.513	13.471	0.5	3.6875	2
95	1.4142	-1.414	-1.414	-1	-1	1	-1	1	-1	-1	26.039	13.941	1.188	3.25	2
96	1.4142	-1.414	-1.414	-1	-1	1	-1	1	-1	-1	18.541	6.444	0.938	3.25	2
97	1.4142	-1.414	-1.414	-1	-1	-1	1	1	-1	1	15.018	3.003	0.5	2.875	2
98	1.4142	1.4142	-1.414	-1	1	-1	1	1	-1	1	79.086	1.288	0	3.125	2
99	1	1	-1.414	-1	1	-1	-1	1	-1	-1	93.808	23.423	0.5	3.5	3
100	-1	-1	-1.414	-1	1	-1	-1	1	-1	-1	36.168	3.535	0	3.5	3
101	1	-1	-1.414	-1	1	-1	-1	1	-1	1	36.339	7.947	0.125	3.5	3
102	1	-1	-1.414	-1	1	-1	-1	-1	-1	-1	30.497	2.112	0	3.3125	3
103	-1	1	-1.414	-1	1	-1	-1	-1	-1	-1	76.416	1.762	0	3.9375	3
104	1	-1	-1.414	-1	1	-1	-1	-1	-1	-1	37.384	8.999	0.563	3.125	3
105	1	-1	-1.414	-1	1	-1	-1	1	-1	1	29.871	1.478	0	3.5625	3
106	1	-1	-1.414	-1	1	-1	-1	1	1	-1	33.715	4.848	0.469	3.4688	3
107	1	1	-1.414	-1	1	-1	-1	-1	1	-1	74.597	6.941	0.625	3.5938	3
108	1	-1	-1.414	-1	1	-1	-1	1	-1	1	29.669	1.277	0	3.4375	3

Raw Data Set (Continued)

OBS#	SR	IT	MUS	AGE	DB	FB	WG	TV	NT	SEX	TST	UAT	AKP	STA	EXP
109		1	-1.414	-1	1	-1	-1	1	1	1	77.387	9.725	0.594	3.5	3
110		1	-1.414	-1	1	-1	-1	-1	1	1	34.464	5.626	0.313	3.3438	3
111		-1	-1.414	-1	1	-1	-1	-1	-1	-1	76.001	1.347	0.063	3.5	3
112		-1	-1.414	-1	1	-1	-1	1	1	1	39.381	5.263	0.531	3.6563	3
113		-1	-1.414	-1	1	-1	-1	-1	1	-1	40.896	6.768	0.281	3.6563	3
114		-1	-1.414	-1	1	-1	-1	1	-1	1	77.184	2.531	0.251	3.5	3
115		-1	-1.414	-1	1	-1	-1	-1	1	1	86.097	13.179	0.688	3.1563	3
116		1	-1.414	-1	1	-1	-1	1	-1	-1	78.895	8.51	0.125	3.5	3
117		1	-1.414	-1	1	-1	-1	-1	-1	1	72.329	1.936	0.063	3.375	3
118		1	-1.414	-1	1	-1	-1	-1	-1	1	79.839	9.446	0.063	3.25	3
119		-1	-1.414	-1	1	-1	-1	1	1	-1	36.173	7.306	0.594	3.4063	3
120		1	-1.414	-1	1	-1	-1	1	-1	-1	72.416	2.031	0.063	3.125	3
121		-1	-1.414	-1	1	-1	-1	-1	1	1	78.813	5.895	0.375	3.8438	3
122		-1	-1.414	-1	1	-1	-1	-1	1	1	77.783	4.865	0.469	3.8438	3
123		-1	-1.414	-1	1	-1	-1	1	1	-1	77.343	4.412	0.563	3.6875	3
124		-1	-1.414	-1	1	-1	-1	-1	1	-1	32.861	-1.268	0.156	3.5625	3
125		-1	-1.414	-1	1	-1	-1	1	-1	1	86.256	11.602	0.25	3.8125	3
126		1	-1.414	-1	1	-1	-1	1	1	1	77.8	10.138	0.626	3.125	3
127		-1	-1.414	-1	1	-1	-1	-1	-1	1	33.94	1.302	0	3.8125	3
128		1	-1.414	-1	1	-1	-1	1	1	-1	32.552	3.685	0.344	3.6875	3
129		1	-1.414	-1	1	-1	-1	1	1	1	69.915	2.253	0.344	3.375	3
130		-1	-1.414	-1	1	-1	-1	1	-1	1	81.699	7.046	0.063	3.5625	3
131		1	-1.414	-1	1	-1	-1	-1	1	-1	76.464	8.808	0.531	3.3125	3
132		-1	-1.414	-1	1	-1	-1	-1	1	-1	38.88	4.751	0.438	3.5	3
133		1	-1.414	-1	1	-1	-1	-1	1	-1	69.821	2.165	0.219	3.4688	3
134		1	-1.414	-1	1	-1	-1	-1	1	1	32.827	3.989	0.438	3.6563	3
135		-1	-1.414	-1	1	-1	-1	1	1	1	37.429	3.311	0.344	3.5625	3

Raw Data Set (Continued)

OBS#	SR	IT	MUS	AGE	DB	FB	WG	TV	NT	SEX	TST	UAT	AKP	STA	EXP #
136		-1	-1	-1.414	-1	1	-1	-1	-1	1	33.783	1.145		0	3.75
137		-1	1	-1.414	-1	1	-1	1	1	-1	78.084	5.154	0.531		3.875
138		-1	1	-1.414	-1	1	-1	-1	-1	-1	76.778	2.125	0.063		3.75
139		1	-1	-1.414	-1	1	-1	-1	1	1	31.853	3.015	0.406		3.4063
140		-1	-1	-1.414	-1	1	-1	-1	-1	1	36.853	4.214		0	3.9375
141		1	1	-1.414	-1	1	-1	-1	-1	1	73.76	3.367	0.313		3.5
142		-1	-1	-1.414	-1	1	-1	1	-1	-1	39.425	6.792	0.626		3.125
143		1	-1	-1.414	-1	1	-1	-1	-1	-1	38.768	10.383	0.688		3.1875
144		-1	1	-1.414	-1	1	-1	1	1	-1	78.467	5.536	0.438		3.6563
145		-1	-1	-1.414	-1	1	-1	1	1	1	39.374	5.257	0.407		3.7813
146		-1	-1	-1.414	-1	1	-1	-1	-1	-1	36.161	3.527		0	3.625
147		-1	1	1	-1	-1	1	1	-1	1	46.782	51.057	0.125		3.375
148		1	-1	1	-1	-1	1	1	-1	1	32.724	-7.686	2.532		1.0938
149		1	1	1	1	1	1	1	-1	-1	74.233	6.537	0.781		2.5313
150		-1	1	1	1	-1	1	-1	1	1	60.129	19.773	1.281		3.0313
151		-1	-1	1	-1	-1	1	-1	-1	1	37.978	15.581	2.75		2.5938
152		1	-1	1	-1	1	1	1	-1	1	31.574	2.671	0.281		2.6563

Appendix C
Reanalysis Results of Screening Study

ANOVA Summary Table for Total Search Time (TST) (Beaudet, 1988)

Source	df	SS	MS	F	p
DB**	1	892.996	892.996	13.42	0.0023*
NT	1	7.892	7.892	0.12	0.7353
SEX	1	1.120	1.120	0.02	0.8985
AGE	1	237.337	237.337	3.57	0.0784*
WG	1	320.513	320.513	4.82	0.0443*
MUS	1	54.680	54.680	0.82	0.3790
SR	1	39.846	39.846	0.60	0.4510
TV	1	8.157	8.157	0.12	0.7311
SRT	1	0.509	0.509	0.01	0.9315
IT	1	1115.762	1115.762	16.77	0.0010*
PS	1	71.910	71.910	1.08	0.3150
SP	1	60.401	60.401	0.91	0.3558
REP	1	0.020	0.020	0.00	0.9865
SF	1	290.393	290.393	4.36	0.0541*
MF	1	60.451	60.451	0.91	0.3556
CF	1	32.526	32.526	0.49	0.4951
Error	15	998.016	66.534		
Total	31	4192.526			

* Significant at $\alpha = 0.20$

** SR=speech rate, IT=input timeout, NT=number of target, TV=type of voice, SEX=sex of subject, AGE=age of subject, WG=wallet guide, MUS=background music, SRT=system response time, DB=database PS=pause/resume, SP=spell-out, REP=repeat, SF=selection feedback, MF=menu feedback, CF=command feedback

ANOVA Summary Table for User Added Time (UAT) (Beaudet, 1988)

Source	df	SS	MS	F	p
DB**	1	32.882	32.882	0.63	0.4408
NT	1	3.333	3.333	0.06	0.8044
SEX	1	1.079	1.079	0.02	0.8878
AGE	1	237.652	237.652	4.53	0.0502*
WG	1	320.754	320.754	6.12	0.0258*
MUS	1	60.055	60.055	1.15	0.3015
SR	1	1.126	1.126	0.02	0.8855
TV	1	4.378	4.378	0.08	0.7766
SRT	1	0.486	0.486	0.01	0.9246
IT	1	34.143	34.143	0.65	0.4323
PS	1	72.000	72.000	1.37	0.2596
SP	1	59.176	59.176	1.13	0.3049
REP	1	0.013	0.013	0.00	0.9875
SF	1	273.605	273.605	5.22	0.0373*
MF	1	61.594	61.594	1.17	0.2956
CF	1	31.776	31.776	0.61	0.4484
Error	15	786.610	52.441		
Total	31	1980.663			

* Significant at $\alpha = 0.20$

** SR=speech rate, IT=input timeout, NT=number of target,TV=type of voice, SEX=sex of subject, AGE=age of subject, WG=wallet guide, MUS=background music, SRT=system response time, DB=database, PS=pause/resume, SP=spell-out, REP=repeat, SF=selection feedback, MF=menu feedback, CF=command feedback

ANOVA Summary Table for User Added Keypresses (AKP) (Beaudet, 1988)

Source	df	SS	MS	F	p
DB**	1	2.709	2.709	4.08	0.0616*
NT	1	3.206	3.206	4.83	0.0441*
SEX	1	0.247	0.247	0.37	0.5509
AGE	1	2.675	2.675	4.03	0.0631*
WG	1	0.317	0.317	0.48	0.5000
MUS	1	1.699	1.699	2.56	0.1305*
SR	1	0.633	0.633	0.95	0.3442
TV	1	0.564	0.564	0.85	0.3711
SRT	1	1.079	1.079	1.63	0.2218
IT	1	0.517	0.517	0.78	0.3914
PS	1	0.598	0.598	0.90	0.3577
SP	1	0.117	0.117	0.18	0.6804
REP	1	0.003	0.003	0.00	0.9466
SF	1	3.327	3.327	5.01	0.0408*
MF	1	0.565	0.565	0.85	0.3709
CF	1	0.185	0.185	0.28	0.6409
<u>Error</u>	15	9.959	9.959		
<u>Total</u>	31	28.401			

* Significant at $\alpha = 0.20$

** SR=speech rate, IT=input timeout, NT=number of target, TV=type of voice, SEX=sex of subject, AGE=age of subject, WG=wallet guide, MUS=background music, SRT=system response time, DB=database, PS=pause/resume, SP=spell-out, REP=repeat, SF=selection feedback, MF=menu feedback, CF=command feedback

ANOVA Summary Table for Transcription Average (STA) (Beaudet, 1988)

Source	df	SS	MS	F	p
DB**	1	0.343	0.343	2.86	0.1116*
NT	1	0.002	0.002	0.02	0.9002
SEX	1	0.548	0.548	4.57	0.0494*
AGE	1	0.668	0.668	5.57	0.0322*
WG	1	0.330	0.330	2.75	0.1179*
MUS	1	0.096	0.096	0.80	0.3858
SR	1	0.841	0.841	7.01	0.0183*
TV	1	0.205	0.205	1.71	0.2106
SRT	1	0.000	0.000	0.00	0.9750
IT	1	0.024	0.024	0.20	0.6615
PS	1	0.141	0.141	1.18	0.2952
SP	1	0.356	0.356	2.97	0.1055*
REP	1	0.343	0.343	2.86	0.1116*
SF	1	0.049	0.049	0.41	0.5331
MF	1	0.581	0.581	4.84	0.0438*
CF	1	0.003	0.003	0.03	0.8754
Error	15	1.799	0.120		
Total	31	6.330			

* Significant at $\alpha = 0.20$

** SR=speech rate, IT=input timeout, NT=number of target, TV=type of voice, SEX=sex of subject, AGE=age of subject, WG=wallet guide, MUS=background music, SRT=system response time, DB=database, PS=pause/resume, SP=spell-out, REP=repeat, SF=selection feedback, MF=menu feedback, CF=command feedback

Appendix D
Reanalysis Results of Previous Studies

ANOVA Summary Table for Total Search Time (TST) (Experiment 1: Wu, 1989)

Source	df	SS	MS	F	p
<u>Regression</u>	(14)	24443.423	1745.959		
SR**	1	414.370	414.370	2.09	
IT	1	20113.000	20113.000	101.61	*
MUS	1	859.945	859.945	4.34	*
AGE	1	1046.626	1046.626	5.29	*
SRxSR	1	16.098	16.098	0.08	
ITxIT	1	0.207	0.207	0.00	
MUSxMUS	1	1.255	1.255	0.00	
AGExAGE	1	263.307	263.307	1.33	
SRxIT	1	96.102	96.102	0.49	
SRxMUS	1	233.005	233.005	1.18	
SRxAGE	1	323.311	323.311	1.63	
ITxMUS	1	950.164	950.164	4.80	*
ITxAGE	1	52.595	52.595	0.27	
MUSxAGE	1	72.960	72.960	0.36	
<u>Residual</u>	(35)	6089.815	173.995		
LOF	10	1141.186	114.119	0.58	
Error	25	4948.629	197.945		
<u>Total</u>	49	30533.238			

* Significant at $\alpha = 0.05$ ($F(1,25)=4.24$; $F(10,25)=2.24$)

** SR=speech rate, IT=input timeout, MUS=background music,
AGE=age of subject

ANOVA Summary Table for User Added Time (UAT) (Experiment 1: Wu, 1989)

Source	df	SS	MS	F	p
<u>Regression</u>	(14)	5136.253	366.875		
SR**	1	1121.567	1121.875	5.57	*
IT	1	110.767	110.767	0.56	
MUS	1	861.015	861.015	4.35	*
AGE	1	1047.809	1047.809	5.29	*
SRxSR	1	5.911	5.911	0.03	
ITxIT	1	0.161	0.161	0.00	
MUSxMUS	1	1.088	1.088	0.00	
AGExAGE	1	260.493	260.493	1.32	
SRxIT	1	96.681	96.681	0.49	
SRxMUS	1	233.010	233.010	1.18	
SRxAGE	1	322.999	322.999	1.63	
ITxMUS	1	949.064	949.064	4.79	*
ITxAGE	1	52.567	52.567	0.27	
MUSxAGE	1	73.030	73.030	0.37	
<u>Residual</u>	(35)	5966.460	170.470		
LOF	10	1.17.843	101.784	0.51	
Error	25	4848.617	197.945		
<u>Total</u>	49	11102.713			

* Significant at $\alpha = 0.05$ ($F(1,25)=4.24$; $F(10,25)=2.24$)

** SR=speech rate, IT=input timeout, MUS=background music,
AGE=age of subject

ANOVA Summary Table for User Added Keypresses (AKP) (Experiment 1: Wu, 1989)

Source	df	SS	MS	F	p
<u>Regression</u>	(14)	7.824	0.559		
SR**	1	1.275	1.275	3.84	
IT	1	0.005	0.005	0.01	
MUS	1	1.809	1.809	5.45	*
AGE	1	1.822	1.822	5.49	*
SRxSR	1	0.124	0.124	0.37	
ITxIT	1	0.148	0.148	0.45	
MUSxMUS	1	0.173	0.173	0.52	
AGExAGE	1	0.229	0.229	0.69	
SRxIT	1	0.001	0.001	0.00	
SRxMUS	1	0.044	0.044	0.12	
SRxAGE	1	0.548	0.548	1.65	
ITxMUS	1	1.196	1.196	3.58	
ITxAGE	1	0.133	0.133	0.39	
MUSxAGE	1	0.317	0.317	0.95	
<u>Residual</u>	(35)	9.479	0.271		
LOF	10	1.184	0.118	0.55	
Error	25	8.295	0.332		
<u>Total</u>	49	17.303			

* Significant at $\alpha = 0.05$ ($F(1,25)=4.24$; $F(10,25)=2.24$)

** SR=speech rate, IT=input timeout, MUS=background music, AGE=age of subject

ANOVA Summary Table for Transcription Average (STA) (Experiment 1: Wu, 1989).

Source	df	SS	MS	F	p
<u>Regression</u>	(14)	10.594	0.757		
SR**	1	1.527	1.527	8.30	*
IT	1	0.008	0.008	0.04	
MUS	1	5.015	5.015	27.26	*
AGE	1	1.031	1.031	5.60	*
SRxSR	1	0.035	0.035	0.19	
ITxIT	1	0.282	0.282	1.52	
MUSxMUS	1	1.196	1.196	6.47	*
AGExAGE	1	0.250	0.250	1.36	
SRxIT	1	0.096	0.096	0.52	
SRxMUS	1	0.305	0.305	1.66	
SRxAGE	1	0.176	0.176	0.96	
ITxMUS	1	0.000	0.000	0.00	
ITxAGE	1	0.040	0.040	0.21	
MUSxAGE	1	0.633	0.633	3.44	
<u>Residual</u>	(35)	7.915	0.226		
LOF	10	3.323	0.332	1.81	
Error	25	4.592	0.184		
<u>Total</u>	49	18.509			

* Significant at $\alpha = 0.05$

** SR=speech rate, IT=input timeout, MUS=background music,
AGE=age of subject

ANOVA Summary Table for Total Search Time (TST) (Experiment 2: Martin, et al., 1990)

Source	df	SS	MS	F	p
<u>Main Effects</u>					
DB**	1	3384.931	3384.931	118.65	0.0001*
WG	1	966.715	966.715	33.89	0.0001*
SR	1	832.217	832.217	29.17	0.0001*
IT	1	22953.528	22953.528	804.57	0.0001*
FB	1	29.116	29.116	1.02	0.3200
<u>Interactions</u>					
DBxWG	1	540.061	540.061	18.93	0.0001*
DBxSR	1	32.321	32.321	1.13	0.2951
DBxIT	1	1970.614	1970.614	69.07	0.0001*
DBxFB	1	90.613	90.613	3.18	0.0842
WGxSR	1	101.617	101.617	3.56	0.0682
WGxIT	1	302.555	302.555	10.61	0.0027*
WGxFB	1	46.701	46.701	1.64	0.2099
SRxIT	1	75.310	75.310	2.64	0.1140
SRxFB	1	81.047	81.047	2.84	0.1016
ITxFB	1	18.379	18.379	0.64	0.4281
Error	32	912.930	28.529		
<u>Total</u>	47	32338.656			

* Significant at $\alpha = 0.05$

** SR=speech rate, IT=input timeout, DB=database structure,
FB=feedback, WG=wallet guide

ANOVA Summary Table for User Added Time (UAT) (Experiment 2: Martin, et al., 1990)

Source	df	SS	MS	F	p
<u>Main Effects</u>					
DB**	1	952.746	952.746	33.40	0.0001*
WG	1	969.014	969.014	33.97	0.0001*
SR	1	94.928	94.928	3.33	0.0775
IT	1	165.897	165.897	5.82	0.0218*
FB	1	27.965	27.965	0.98	0.3296
<u>Interactions</u>					
DBxWG	1	537.823	537.823	18.85	0.0001*
DBxSR	1	31.179	31.179	1.09	0.3037
DBxIT	1	121.680	121.680	4.27	0.0471*
DBxFB	1	86.570	86.570	3.03	0.0911
WGxSR	1	103.758	103.758	3.64	0.0655
WGxIT	1	301.652	301.652	10.57	0.0027*
WGxFB	1	47.299	47.299	1.66	0.2071
SRxIT	1	74.053	74.053	2.60	0.1170
SRxFB	1	79.779	79.779	2.80	0.1042
ITxFB	1	17.880	17.880	0.63	0.4344
<u>Error</u>	32	912.898	28.528		
<u>Total</u>	47	4525.122			

* Significant at $\alpha = 0.05$

** SR=speech rate, IT=input timeout, DB=database structure,
FB=feedback, WG=wallet guide

ANOVA Summary Table for User Added Keypresses (AKP) (Experiment 2:
Martin, et al., 1990)

Source	df	SS	MS	F	p
<u>Main Effects</u>					
DB**	1	4.649	4.649	14.21	0.0007*
WG	1	4.122	4.122	12.60	0.0012*
SR	1	0.078	0.078	0.24	0.6289
IT	1	0.228	0.228	0.70	0.4098
FB	1	0.036	0.036	0.11	0.7424
<u>Interactions</u>					
DBxWG	1	2.967	2.967	9.07	0.0050*
DBxSR	1	0.069	0.069	0.21	0.6503
DBxIT	1	0.507	0.507	1.55	0.2222
DBxFB	1	0.010	0.010	0.03	0.8637
WGxSR	1	0.457	0.457	1.40	0.2458
WGxIT	1	1.529	1.529	4.67	0.0383*
WGxFB	1	0.069	0.069	0.21	0.6503
SRxIT	1	0.561	0.561	1.71	0.1997
SRxFB	1	0.560	0.560	1.71	0.2000
ITxFB	1	0.007	0.007	0.02	0.8881
<u>Error</u>	32	10.472	0.327		
<u>Total</u>	47	26.320			

* Significant at $\alpha = 0.05$

** SR=speech rate, IT=input timeout, DB=database structure,
FB=feedback, WG=wallet guide

ANOVA Summary Table for Transcription Average (STA) (Experiment 2:
Martin, et al., 1990)

Source	df	SS	MS	F	p
<u>Main Effects</u>					
DB**	1	0.014	0.014	0.21	0.6477
WG	1	0.002	0.002	0.03	0.8603
SR	1	1.572	1.572	24.33	0.0001*
IT	1	0.001	0.001	0.01	0.9159
FB	1	0.007	0.007	0.10	0.7515
<u>Interactions</u>					
DBxWG	1	0.007	0.007	0.10	0.7515
DBxSR	1	0.018	0.018	0.28	0.5982
DBxIT	1	0.004	0.004	0.06	0.8054
DBxFB	1	0.165	0.165	2.55	0.1201
WGxSR	1	0.014	0.014	0.21	0.6477
WGxIT	1	0.007	0.007	0.10	0.7515
WGxFB	1	0.043	0.043	0.67	0.4204
SRxIT	1	0.111	0.111	1.72	0.1985
SRxFB	1	0.036	0.036	0.56	0.4616
ITxFB	1	0.124	0.124	1.92	0.1759
<u>Error</u>	32	2.068	0.065		
<u>Total</u>	47	4.191			

* Significant at $\alpha = 0.05$

** SR=speech rate, IT=input timeout, DB=database structure,
FB=feedback, WG=wallet guide

Appendix E
Informed Consent Form

Participant's Informed Consent

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 indepth 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 Robert C. Williges, and Beverly H. Williges. The researcher is Sung H. Han. 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 Dr. Stout, 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:

Dr. E. R. Stout
Office of Sponsored Research Programs
306 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 your name and address
if you wish to receive
experimental results

Appendix F
Pre-test 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 G
Subject's Instructions

INTRODUCTION

Hello, and welcome to the Human-Computer Interaction Lab. Today, you have the opportunity to participate in our research on how people interact with talking computers.

In this experiment, you will try to find information on certain items in a department store named Hokie Wholesale. The department store has a talking computer database system which provides shoppers with helpful information on store items. Shoppers call the database system on a telephone to find information on selected merchandise. Similarly, you will be using the telephone to find specific information in the database. The talking computer may sound a bit strange at first, but we are sure you will soon be able to understand everything it says. The computer does not understand human speech, but does interpret certain key presses on the telephone keypad as commands.

The database system works by speaking menus of keywords. Keywords are titles for a group of related items (e.g. automotive is a keyword for a group of items like tires, car batteries, and motor oil). When you hear a keyword which most closely relates to the item you are searching for, select that keyword by pressing a defined key on the telephone keypad. The system will then speak a new menu of keywords related to the selected keyword. By selecting the appropriate keywords, you locate the store item in the database. Once you have selected the store item, the computer will speak a short information message about the store item. This message will have something to do with the price, location, availability, or important information about the store item.

INSTRUCTIONS

(Single Task)

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 dial tone.
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.

INSTRUCTIONS

(Multiple Tasks)

Your task is to search for information on store items in the department store's talking database. Two store items will be presented as a target on the computer display in front of you. You will find each store item 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 dial tone.
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. The target requests that you find the information message for two store items.
5. Watch the computer display. It will signal you when the search is about to begin. Begin by searching for the first store item.
6. The talking computer will begin speaking a menu of keywords.

Keywords categorize a group 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 first store item, 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 on the computer keyboard.
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. Review the target on the computer display and get ready to start the search for the second store item. The computer display will signal you to begin the search and will speak the information message for the first store item.
16. Search for the second store item using the same techniques which were described in steps 7 through 10. **Remember**, you are starting the search from the information message for the first store item, not from the main menu.
17. Locate the second store item and transcribe the information message.
18. 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 each store item and transcribe their respective information messages.

19. The experiment will proceed in this fashion. You will search for a total of 16 targets.
20. The computer display will indicate when you have completed the target searches. The computer will then request that you rate certain characteristics of the telephone information system. The meaning of each rating and how it should be rated will be explained on the computer display.

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

INSTRUCTIONS

(Commands)

The video instructions you just watched included a demonstration of how the telephone information system works and how you should perform the task for this study. The actual telephone information system you will be using today will be similar to the system in the video, but may be different in some ways.

These are the commands that are available to you on the telephone keypad:

To select an item, press the **# key**.

To back-up one menu, press the *** key**.

To select the main menu, press the **0 key**.

When you are ready to transcribe the information message, press the
2 key.

Appendix H

Targets and Information Messages

Store Items for Single Target Searches

The following are the single target 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)

Store Items for the Multiple Target Searches

The following are the first store items and associated information messages for the multiple targets. The second store item are the single store items presented above. Information messages are classified as being an availability (A), Information (I), location (L), or price (P) oriented messages.

Love Seat Couches: Many patterns are on the west wall. (L)

Water Beds: Merchandise exchanges are required within 1 month. (I)

Refrigerators: Large freezers are sold for \$500.65. (P)

Telephones: Mouth pieces are available at the cash register. (A)

Stereo Turntables: Rubber belts are sold for \$6.49. (P)

Trumpets: Free demonstrations are offered at most locations. (I)

Candy Cookbooks: Holiday specials are available in hardcover and paperback. (A)

Sailing Books: Instruction manuals are near the road maps. (L)

Men's Summer Suits: Seer sucker cloth is available in 4 colors. (A)

Men's Sweatpants: Running gear is near the camping tents. (L)

Cocktail Dresses: Colonial styles are reduced by 35 to 45%. (P)

Straight Skirts: Spring mini-skirts are offered with matching pocketbooks. (I)

Silver Cufflinks: Antique objects are sold for \$64.99. (P)

Diamond Pins: Custom orders are offered to graduate students. (I)

Nail Polish: Applicator brushes are near the hair shampoo. (L)

Musk Fragrances: Bottled after-shave is available with rebate coupons. (A)

Appendix I
After-test Questionnaire

Subject Debrief

Subject No. _____

1. Have you used an **information system** like this one, before?

YES _____

NO _____

Would you use one? YES _____ NO _____

2. Which **system features** do you like/dislike?

Like:

Dislike:

3. Were the **instructions** sufficient enough?

YES _____

NO _____

Are there any **instructions** that you would like to add/delete?

Add:

Delete:

4. Do you understand the phone **key commands**?

YES _____

NO _____

5. Are there any **commands** that you would like to add/delete?

Add:

Delete:

6. What do you think of the **database organization**?

Was it confusing?

YES____

NO____

7. Was the **speech rate** appropriate?

YES____

NO____

8. **Input time-out:** Were you given **too much / not enough** time to respond to the computer?

Appendix J
Theoretical Background for Data Bridging

Theoretical Background for Data Bridging

Suppose there are ten variables ($X_1, X_2, \dots, X_{10}$) to be investigated in the description stage of a sequential research plan, and X_1 is selected as a system parameter. The variables, X_1, X_2, X_3 , and X_4 are continuous, and it is required to investigate second-order effects of the continuous variables. The other six variables ($X_5, X_6, \dots, X_{10}$) are dichotomous. The experimenter divides the set of ten independent variables into three groups; Group 1: X_1, X_2, X_3 , and x_4 ; Group 2: X_1, X_5 , and X_6 ; and Group 3: X_1, X_7, X_8, X_9 , and X_{10} . A CCD is used to investigate the variables in Group 1; whereas a 2^3 factorial design is used to investigate the variables in Group 2 and a 2^{5-1} fractional factorial design is used for the variables in Group 3. Further, the experimental region of interest is assumed to be spherical for the continuous variables (i.e., $\alpha > 1.0$ for the CCD). Table M-1, M-2, and M-3 show the design matrices of the three experiments using coded values.

Identifying uninvestigated interactions. To investigate a two-factor interaction, at least four different design points are required as long as each level of a factor is related to at least two different levels of the other factor. That is, it is impossible to estimate a two-factor interaction effect with less than four different design points and even with four design points if more than two different design points (i.e., three or four) are on the same line within the experimental space created by the two factors. Figure M-1 illustrates some examples where it is impossible to estimate the two-factor interaction effect between X_1 and X_2 .

TABLE M-1
Design Matrix for a Four-Factor, Second-Order, Central-
Composite Design for Variables X_1 , X_2 , X_3 , and X_4 in Group 1.

X_1	X_2	X_3	X_4
+1	+1	+1	+1
-1	-1	-1	-1
-1	-1	-1	+1
-1	-1	+1	-1
-1	+1	-1	-1
+1	-1	-1	-1
-1	-1	+1	+1
-1	+1	-1	+1
+1	-1	-1	+1
-1	+1	+1	-1
+1	-1	+1	-1
+1	+1	-1	-1
-1	+1	+1	+1
+1	-1	+1	+1
+1	+1	-1	+1
+1	+1	+1	-1
$-\alpha$	0	0	0
$+\alpha$	0	0	0
0	$-\alpha$	0	0
0	$+\alpha$	0	0
0	0	$-\alpha$	0
0	0	$+\alpha$	0
0	0	0	$-\alpha$
0	0	0	$+\alpha$
0	0	0	0

Estimable Effects: X_1 , X_2 , X_3 , X_4 , X_1X_2 , X_1X_3 , X_1X_4 , X_2X_3 , X_2X_4 , X_3X_4 , X_1^2 , X_2^2 , X_3^2 , and X_4^2

Note: Variables X_5 , X_6 , X_7 , X_8 , X_9 , and X_{10} are maintained at the +1 coded value.

TABLE M-2
 Design Matrix for a 2³ Factorial Design for Variables X₁, X₅,
 and X₆ in Group 2.

X ₁	X ₅	X ₆
+1	+1	+1
-1	-1	-1
-1	-1	+1
-1	+1	-1
+1	-1	-1
-1	+1	+1
+1	-1	+1
+1	+1	-1

Estimable Effects: X₁, X₅, X₆, X₁X₅, X₁X₆, X₅X₆

Note: Variables X₂, X₃, X₄, X₇, X₈, X₉, and X₁₀ are maintained at the +1 coded value.

TABLE M-3

Design Matrix for a 2^{5-1} Fractional-Factorial for Variables X_1 , X_7 , X_8 , X_9 , and X_{10} in Group 3.

X_1	X_7	X_8	X_9	X_{10}
+1	+1	+1	+1	+1
+1	+1	+1	-1	-1
+1	+1	-1	+1	-1
+1	-1	+1	+1	-1
-1	+1	+1	+1	-1
+1	+1	-1	-1	+1
+1	-1	+1	-1	+1
-1	+1	+1	-1	+1
+1	-1	-1	+1	+1
-1	+1	-1	+1	+1
-1	-1	+1	+1	+1
-1	-1	-1	-1	+1
-1	-1	-1	+1	-1
-1	-1	+1	-1	-1
-1	+1	-1	-1	-1
+1	-1	-1	-1	-1

Estimable Effects: X_1 , X_7 , X_8 , X_9 , X_{10} , X_1X_7 , X_1X_8 , X_1X_9 , X_1X_{10} , X_7X_8 , X_7X_9 , X_7X_{10} , X_8X_9 , X_8X_{10} , X_9X_{10}

Note: Variables X_2 , X_3 , X_4 , X_5 , and X_6 are maintained at the +1 coded value.

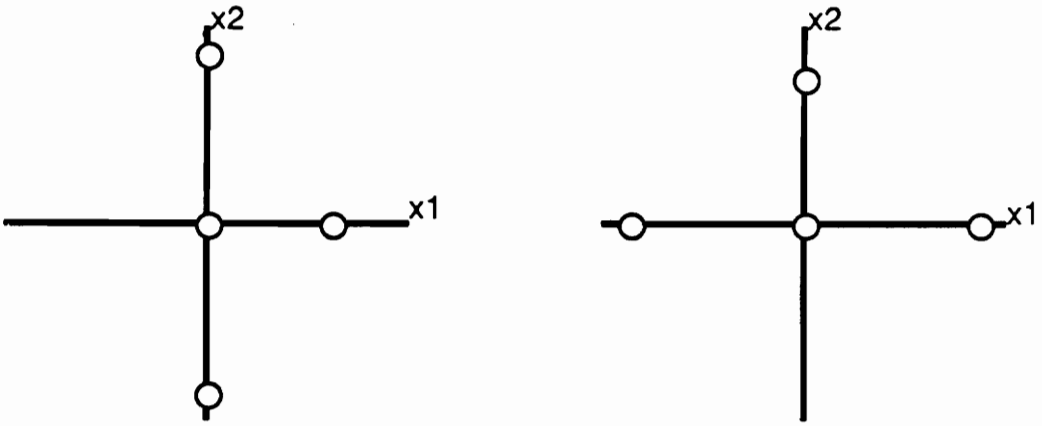


Figure M-1. Some examples where it is impossible to investigate a two-factor interaction with four design points.

To find additional design points required to build a second-order empirical model, the experimenter must determine the experimental region of interest. When each experimental variable is defined by an upper and lower bound, the experimental region is a hypercube (i.e., a cuboidal region). If experimentation in the corners of the hypercube is impossible or unnecessary, it may be appropriate to employ a spherical design in which a hypersphere, cutting off the corners of the hypercube, is used. For a spherical region of interest, additional design points should be determined in such a way that they are located inside a hypersphere generated by the experimental variables. For example, if the experimenter used a CCD for a spherical region of interest (i.e., $\alpha > 1.0$), then an additional design point should be selected within the region generated by the factorial portion in the CCD since the information required to estimate interaction effects is derived from the factorial portion.

From the design matrices in Table M-1, M-2, and M-3, it is easily found that the following 26 interactions cannot be estimated without additional design point even though the data sets resulting from the three studies are combined together into one integrated data set. These uninvestigated interactions include: X_2X_5 , X_2X_6 , X_2X_7 , X_2X_8 , X_2X_9 , X_2X_{10} , X_3X_5 , X_3X_6 , X_3X_7 , X_3X_8 , X_3X_9 , X_3X_{10} , X_4X_5 , X_4X_6 , X_4X_7 , X_4X_8 , X_4X_9 , X_4X_{10} , X_5X_7 , X_5X_8 , X_5X_9 , X_5X_{10} , X_6X_7 , X_6X_8 , X_6X_9 , and X_6X_{10} .

Selecting additional design points. For each uninvestigated interaction, it is necessary to provide one additional design point to provide four different combinations of design points required to estimate the interaction since the combined data set provides three different design points for each uninvestigated interaction. In other words, 26 additional design points should

be added to the combined data set resulting from the three studies to estimate the 26 uninvestigated interactions. For example, the uninvestigated interaction between X_2 and X_4 (i.e., X_2X_4) has three different design points in the combined data set. The three design points are $(X_2, X_4) = (1, 1)$, $(-1, 1)$, and $(1, -1)$. Therefore, only one additional design point needs to be added to provide four different design points required to estimate the interaction between X_2 and X_4 .

In selecting an additional design point, it is necessary to develop decision criteria to obtain desirable statistical properties such as prediction variances and variances of coefficients since there is a great deal of possible candidates for the fourth design point. For example, in selecting the fourth design point for the investigation of X_2X_4 , an arbitrary number between -1 and +1 can be assigned to the level of X_2 since X_2 is a continuous variable. It is generally known, however, that the variance of coefficient of a model item in a linear regression equation is minimum when the factor levels are set at its extremes within its region of interest (Dykstra, 1971). Therefore, $(X_2, X_4) = (-1, -1)$ provides the optimal additional design point among the infinite number of candidates since the information for estimating an interaction effect is stemming from the factorial portion rather than the axial or center points in a spherical CCD.

Likewise, the fourth design points of all uninvestigated interactions can be determined based on the criterion described above. Table M-4 shows the fourth design points for all uninvestigated interactions in the simulation study.

As shown in Table M-4, the additional design point for each uninvestigated interaction can be expressed as the levels between the two

TABLE M-4

The Fourth Design Point for the Uninvestigated Interactions in the Simulation Study.

Uninvestigated Interaction	Factor				Level					
	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	Δ_9	Δ_{10}
X2X5		-1			-1					
X2X6		-1				-1				
X2X7		-1					-1			
X2X8		-1						-1		
X2X9		-1							-1	
X2X10		-1								-1
X3X5			-1		-1					
X3X6			-1			-1				
X3X7			-1				-1			
X3X8			-1					-1		
X3X9			-1						-1	
X3X10			-1							-1
X4X5				-1	-1					
X4X6				-1		-1				
X4X7				-1			-1			
X4X8				-1				-1		
X4X9				-1					-1	
X4X10				-1						-1
X5X7					-1		-1			
X5X8					-1			-1		
X5X9					-1				-1	
X5X10					-1					-1
X6X7						-1	-1			
X6X8						-1		-1		
X6X9						-1			-1	
X6X10						-1				-1

factors. However, Table M-4 does not show how a specific treatment combination of all ten factors should be selected as a final design point. For example, the fourth design point for X_2X_4 can be any one of 2^8 combinations generated by the other factors (i.e., $X_1, X_3, X_5, X_6, X_7, X_8, X_9$, and X_{10}). Therefore, decision criteria for selecting the fourth design point should be determined in order to choose the best design point among the 2^8 candidates.

Statistical criteria for selecting additional design points. When the experimenter wants to build an empirical model for describing the functional relationships between independent variables and their responses, the following linear model approximation is usually used (Myers, 1990);

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations, $\mathbf{X}$ is an $n \times k$ of constants, $\boldsymbol{\beta}$ is the $k \times 1$ vector of coefficients to be estimated, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random variables. There are a variety of statistical criteria for constructing optimal designs to fit the linear model. One of the most popular is the maximization of $|\mathbf{X}'\mathbf{X}|$, i.e., minimization of the generalized variance of the estimates of the coefficients (Mitchell, 1974a). One of the attractive features of the $|\mathbf{X}'\mathbf{X}|$ criterion is that designs which are optimal with respect to it are invariably good in many respects (e.g., low variances for the coefficients, low correlations among the coefficients, and low maximum prediction variance over the experimental region of interest) (Mitchell, 1974b).

In augmenting existing data with additional design points, the $|\mathbf{X}'\mathbf{X}|$ criterion has been used to find appropriate design points (Dykstra, 1971; Evans, 1979; Hebble and Mitchell, 1972; Mitchell, 1974a, 1974b). This criterion is based on a theory that describes the relationship between the maximization of

$|X'X|$ and prediction variance within the design region of interest. That is, the additional design points should be selected at the point(s) where the prediction variance is maximum. Dykstra (1971) proved that the design point can be obtained by maximizing the $|X'X|$. However, this procedure requires the experimenter to specify an initial set of candidates for the additional design points. In choosing specific combinations of the candidates, the experimenter should be guided by the model which he wants to build (Dykstra, 1971). For a first-order model candidates should be selected at the extremes of the experimental space, so that only corner points are specified as candidates. For a second-order model the list of candidates should include axial points and a center point, in addition to corner points.

Appendix K
Computer Program List

Computer Program for Calculating |X'X|

DETX PRT A1 01/14/91 1:09 V 72 226 RECS 01/25/91 13:01 PAGE

```

    INTEGER INTR(21,10),ESBI(5),ESBJ(5),NESBI(21),
+       NESBJ(21),NINV,NUNI,IFLAG(26),DORDER(1024),IPVT(15),N2,
+       MEIDX(7),QEIDX(1)
    REAL DEXT(200,10),XMTX(200,40),RCAND(1024,10),
+DET1,DET2,DETVAL(1024),DRP(21,146,2),XPRX(15,15),
+FAC(15,15),AINV(15,15),ADDDATA(10,10)
    DOUBLE PRECISION EVAL(15),A(15,15),RATILAM
    EXTERNAL LFTRG,LFDRG,DLINRG,DEVLSF,WRRRN
    NVAR=10
    NOBS=146
    NCAND=1024
    NQUAD=1
    NINTR=21
    NINV=5
    NUNI=16
*****
*   READ EXISTING DATA FILE
*****
    READ(11,800) ((DEXT(I,J),J=1,NVAR),I=1,NOBS)
  800  FORMAT(10(F8.4))
    CALL RDINTR(INTR,NINTR,NVAR,NINV,NUNI,ESBI,ESBJ,NESBI,NESBJ,IFLAG)
    CALL RDCAND(RCAND,NCAND,NVAR)
*****
*   BUILD MODEL WITH EXISTING DATA
*****
    NMAIN=7
    MEIDX(1)=1
    MEIDX(2)=2
    MEIDX(3)=3
    MEIDX(4)=4
    MEIDX(5)=5
    MEIDX(6)=7
    MEIDX(7)=9
    DO 51 I=1,NOBS
      XMTX(I,1)=1.0
      DO 51 J=2,NMAIN+1
51    XMTX(I,J)=DEXT(I,MEIDX(J-1))
      DO 61 I=1,NOBS
        DO 31 J=NMAIN+2,NMAIN+NINV+1
31    XMTX(I,J)=DEXT(I,ESBI(J-NMAIN-1))*DEXT(I,ESBJ(J-NMAIN-1))
      IF (NQUAD.EQ.0) THEN
        GO TO 61
      END IF
      XMTX(I,NMAIN+NINV+NQUAD+1)=DEXT(I,3)**2
61    CONTINUE
*****
*   ELIMINATING EXISTING DESIGN POINTS
*****
    DO 1002 I=1,NUNI
      DO 1003 K=1,NOBS
        DRP(I,K,1)=DEXT(K,NESBI(I))
        DRP(I,K,2)=DEXT(K,NESBJ(I))
1003  CONTINUE
1002  CONTINUE
    KCNTR=NOBS

```

```

*****
* TO ADD AN ADDITIONAL DATA POINT TO THE EXISTING MODEL
*****
      NADD=5
      READ(14,1076) ((ADDDATA(IR,JR),JR=1,NVAR),IR=1,NADD)
1076  FORMAT(10(1X,F4.1))
      DO 1077 IA=NOBS+1,NOBS+NADD
      XMTX(IA,1)=1.0
      DO 1021 JA=2,NMAIN+1
      XMTX(IA,JA)=ADDDATA(IA-NOBS,MEIDX(JA-1))
1021  CONTINUE
      DO 1031 JB=NMAIN+2,NMAIN+NINV+1
      XMTX(IA,JB)=ADDDATA(IA-NOBS,ESBI(JB-NMAIN-1))*
+ADDDATA(IA-NOBS,ESBJ(JB-NMAIN-1))
1031  CONTINUE
      IF (NQUAD.EQ.0) THEN
      GO TO 1077
      ELSE
      XMTX(IA,NMAIN+NINV+NQUAD+1)=ADDDATA(IA-NOBS,3)**2
      END IF
1077  CONTINUE
*****
* CALCULATING X'X MATRIX
*****
10001 DO 10000 NU=1,NUNI
      NMODIT=NMAIN+NINV+NQUAD+2
      KWR=0
2002  DO 4000 LL=1,NOBS
      XMTX(LL,NMODIT)=DEXT(LL,NESBI(NU))*DEXT(LL,NESBJ(NU))
4000  CONTINUE
      IF (NADD.EQ.0) THEN
      GO TO 2999
      END IF
      DO 4001 LL=NOBS+1,NOBS+NADD
      XMTX(LL,NMODIT)=ADDDATA(LL-NOBS,NESBI(NU))*
+ADDDATA(LL-NOBS,NESBJ(NU))
4001  CONTINUE
2999  DO 2001 K=1,KCNTR
      IF (RCAND(I,NESBI(NU)).EQ.DRP(NU,K,1)) THEN
      IF (RCAND(I,NESBJ(NU)).EQ.DRP(NU,K,2)) THEN
      DETVAL(I)=0.0
      KWR=KWR+1
      GO TO 5999
      END IF
      END IF
2001  CONTINUE
      XMTX(NOBS+NADD+1,1)=1.0
      DO 5001 J=2,NMAIN+1
      XMTX(NOBS+NADD+1,J)=RCAND(I,MEIDX(J-1))
5001  CONTINUE
      DO 5002 J=NMAIN+2,NMAIN+NINV+1
      XMTX(NOBS+NADD+1,J)=RCAND(I,ESBI(J-NMAIN-1))*
+RCAND(I,ESBJ(J-NMAIN-1))
5002  CONTINUE
      IF (NQUAD.EQ.0) THEN

```

```

GO TO 100
END IF
XMTX(NOBS+NADD+1,NMAIN+NINV+NQUAD+1)=RCAND(I,3)**2
100  NMODIT=NMAIN+NINV+NQUAD+2
XMTX(NOBS+NADD+1,NMODIT)=RCAND(I,NESBI(NU))*RCAND(I,NESBJ(NU))
NOBSU=NOBS+NADD+1
193  DO 202 I1=1,NMODIT
DO 202 J1=1,NMODIT
XPRX(I1,J1)=0.0
202  CONTINUE
DO 104 II=1,NMODIT
DO 104 JJ=1,NMODIT
DO 104 KK=1,NOBSU
XPRX(II,JJ)=XPRX(II,JJ)+XMTX(KK,II)*XMTX(KK,JJ)
104  CONTINUE
*****
* CALCULATING DETERMINANT VALUE
*****
LDA=NMODIT
CALL LFTRG(NMODIT,XPRX,LDA,FAC,LDA,IPVT)
CALL LFDGR(NMODIT,FAC,LDA,IPVT,DET1,DET2)
DETVAL(I)=DET1*10**DET2
5999 DORDER(I)=I
5000 CONTINUE
DO 6000 J=1,NCAND
K=NCAND-J
DO 6001 I=1,K
IF (DETVAL(I).GE.DETVAL(I+1)) THEN
GO TO 6001
END IF
XMAX=DETVAL(I)
DETVAL(I)=DETVAL(I+1)
DETVAL(I+1)=XMAX
IMAXID=DORDER(I)
DORDER(I)=DORDER(I+1)
DORDER(I+1)=IMAXID
6001 CONTINUE
6000 CONTINUE
WRITE(*,243) NESBI(NU),NESBJ(NU)
243  FORMAT(/,'UNINVESTIGATED INTERACTION = X',I2,'X',I2,/)
WRITE(*,111)
111  FORMAT(6X,'X1',4X,'X2',4X,'X3',4X,'X4',4X,'X5',4X,
+'X6',4X,'X7',4X,'X8',4X,'X9',4X,'X10',
+6X,'DET VALUE',3X,'COND',/)
DO 112 I=1,NCAND-KWR
112  WRITE(*,113) (RCAND(DORDER(I),J), J=1,NVAR), DETVAL(I), DORDER(I)
113  FORMAT(3X,10(2X,F4.1),3X,E12.4,3X,I4)
10000 CONTINUE
9999 STOP
END
SUBROUTINE RDINTR(INTR,NINTR,NVAR,NINV,NUNI,ESBI,ESBJ,
+NESBI,NESBJ,IFLAG)
*****
* THIS SUBROUTINE READS INDEXES OF INTERACTIONS AND CLASSIFIES THEM

```

```

* INTO TWO GROUPS: INVESTIGATED AND UNINVESTIGATED INTERACTION.
* INTR; INTERACTIONS AMONG THE NVAR FACTORS
*   ESB; FIRST INDEX OF INTERACTION (INVESTIGATED)
*   ESB; SECOND INDEX OF INTERACTION (INVESTIGATED)
*   NESBI; FIRST INDEX OF INTERACTION (UNINVESTIGATED)
*   NESBJ; SECOND INDEX OF INTERACTION (UNINVESTIGATED)
*   (EX) ESB(1)=1 ESB(2)=2; INTERACTION BETWEEN VARIABLE1 AND
*         VARIABLE2 , IFLAG=1 MEANS INVESTIGATED INTERACTION
*         NESBI(25)=3 NESBJ(25)=5; INTERACTION BETWEEN VARIABLE3
*         AND VARIABLE5, IFLAG=0 MEANS UNINVESTIGATED INTERACTION
*   NIN; NUMBER OF INVESTIGATED INTERACTION
*   NUNI; NUMBER OF UNINVESTIGATED INTERACTION
*
*****
      INTEGER INTR(NINTR,NVAR),ESBI(NINV),ESBJ(NINV),
      +NESBI(NUNI),NESBJ(NUNI),IFLAG(NINTR)
      DO 10 I=1,NINTR
      READ(12,1001) INTR(I,1),INTR(I,2),INTR(I,3),INTR(I,4),
      +INTR(I,5),INTR(I,6),INTR(I,7),INTR(I,8),INTR(I,9),INTR(I,10),
      +IFLAG(I)
10      CONTINUE
1001    FORMAT(11(1X,I1))
      DO 29 L=1,NINTR
      DO 30 J=1,NVAR-1
      DO 30 K=J+1,NVAR
      NCOM=INTR(L,J)*INTR(L,K)
      IF (NCOM.EQ.0) THEN
        GO TO 30
      ELSE
        IF (IFLAG(L).EQ.1) THEN
          ESB(L)=J
          ESB(L)=K
          GO TO 29
        ELSE
          NESBI(L-NINV)=J
          NESBJ(L-NINV)=K
          GO TO 29
        ENDIF
      END IF
30      CONTINUE
29      CONTINUE
85      CONTINUE
      RETURN
      END
      SUBROUTINE RDCAND(RCAND,NCAND,NVAR)
*****
* THIS SUBROUTINE READS CANDIDATE DESIGN POINTS FOR UNINVESTIGATED
* INTERACTIONS.
*   CAND; RAW DATA FROM THE INPUT FILE
*   RCAND; REAL VALUE OF CAND (SINCE THE DATA FILE IS CREATED USING
*         INTEGER NUMBERS
*   NCAND; NUMBER OF CANDIDATES
*   NVAR; NUMBER OF VARIABLES
*

```

DETX PRT A1 01/14/91 1:09 V 72 226 RECS 01/25/91 13:01 PAGE

```
*****  
      REAL RCAND(NCAND,NVAR)  
      READ(13,1000) ((RCAND(I,J),J=1,NVAR),I=1,NCAND)  
1000  FORMAT(10(2X,F4.1))  
      RETURN  
      END
```

VITA

Sung Ho Han

Home:

R176 Shawnee Apt.
Blacksburg, VA 24060

(703) 961-9387

Office:

Industrial & Systems Engineering
302 Whittemore Hall
Virginia Poly. Inst. & State U.
Blacksburg, VA 24061

(703) 231-4602

Education

Ph.D. Industrial and Systems Engineering (Jan. 1991)

Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
Dissertation: " Integrated empirical models based on a sequential
research strategy ". Advisor: Dr. Robert C. Williges.
QCA: 3.92/4.0

M.S. Industrial Engineering (Feb. 1985)

Seoul National University, Seoul, Korea
Thesis: "Effects of a lumbar support device of a driver seat on back
muscle fatigue "
QCA: 3.65/4.3

B.S. Industrial Engineering (Feb. 1983)

Seoul National University, Seoul, Korea
QCA: 3.15/4.3

Professional & Research Appointments

Human-Computer Interaction Lab. (May, 1989 - Jan. 1991)

Department of Industrial & Systems Engineering
VPI&SU

Graduate Research Assistant.

Research topics: Building an empirical model for optimizing the usability of a
telephone-based information system.
Developing a new research paradigm for human factors
research.

Management Systems Lab. (Sep. 1988 - Dec. 1988)
Department of Industrial & Systems Engineering
VPI&SU

System Administrator: Responsible for installing, administering, and maintaining the PROFS software application for an IBM 9370 computer system as well as developing and training software applications in a laboratory environment.

IBM, Korea (Jan. 1986 - Jul. 1987)
Public Branch
Seoul, Korea

System Engineer: In charge of installing and maintaining system and application softwares for IBM mainframes in academic environments. Also responsible for training system programmers to develop software applications.

Human Factors Engineering Lab. (Mar. 1985 - Dec. 1985)
Department of Industrial Engineering
Seoul National University

Research Associate: Involved in two research projects for creating an expert system for automobile seat design; and building a biomechanical model for weight-lifting. Responsible for planning, scheduling, and conducting research as a project team manager.

Human Factors Engineering Lab. (Jul. 1981 - Feb. 1985)
Department of Industrial Engineering
Seoul National University

Research Assistant: Involved in several research projects such as developing a decision support system for a textile company (1981-1982); constructing a master plan for developing a decision support system (1982-1983); creating an expert system for automobile seat design (1983-1985); and building a biomechanical model for weight-lifting (1984-1985).

Teaching Appointments

Graduate Teaching Assistant (Jan. 1989 - Apr. 1989)
Department of Industrial & Systems Engineering
VPI&SU
Course: Human Factors Research Methodology II.

Responsible for preparing and presenting course lectures for SAS programming as well as teaching, preparing, and grading all tests and homework assignments.

Graduate Teaching Assistant (Mar. 1983 - Feb. 1985)

Department of Industrial Engineering

Seoul National University

Courses: Human Factors Engineering, Human Factors Engineering Lab.,
Safety Management, Introduction to Management System

Responsible for preparing and teaching an entire course lecture (Human Factors Engineering Lab.), and administering and grading tests and homework assignments.

Professional Affiliations

Human Factors Society

Association for Computing Machinery:

Computer-Human Interaction Special Interest Group (SIGCHI)

Artificial Intelligence Special Interest Group (SIGART)

Computer Graphics Special Interest Group (SIGGRAPH)

Korean Scientists and Engineers Association in America



Sung Ho Han

Birthdate: Jan. 3, 1962