THE EFFECTS OF SYSTEM RESPONSE TIME AND COGNITIVE LOADING ON ACCESSING AN AUTOMATED TELEPHONE EMERGENCY SERVICE: EXAMINING ELDERLY AND YOUNG USERS

by

Jonathan K. Kies

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

MASTER OF SCIENCE
in
Industrial and Systems Engineering

Approved:

Dr. Dennis Price, Chairman

Dr. Robert Williges

Dr. Robert Springer

November 15, 1994
Blacksburg, Virginia

THE EFFECTS OF SYSTEM RESPONSE TIME AND COGNITIVE LOADING ON ACCESSING AN AUTOMATED TELEPHONE EMERGENCY SERVICE: EXAMINING ELDERLY AND YOUNG USERS

by

Jonathan K. Kies

Chairman: Dr. Dennis L. Price
Department of Industrial and Systems Engineering

(ABSTRACT)

The user interface for a proposed alternative emergency service was conceived and designed for a standard touch-tone telephone. The service would allow a user to activate an automated, pre-recorded message containing information to aid emergency responders. The user must only press a few specified keys on the telephone key-pad, avoiding the need for verbal interaction with a dispatcher.

The interface was designed in terms of providing the necessary instructions for activation and considering various input strategies and feedback. Icons, written instructions, and voice feedback were employed in the development of a successful and effective interface between the user and the system.

Because the system is expected to attract elderly users and families with young children, the performance and attitudes of these two age groups in regard to a system prototype were examined to determine if the interface was suitable.

A two and eight second initial system response time were imposed upon users to determine any effect these delays might have on user response time, error rate, and subjective attitudes. Additionally, a secondary task , designed to increase cognitive loading was employed to determine if the system is usable while the user is engaged in a

dual-task environment. The dependent variables used to gauge the effects of the manipulated variables include the objective measures of user response time and error rate and subjective questionnaire responses.

The results of the study indicate that the elderly adults and young children were able to activate the system successfully. System response time and cognitive loading had no significant effect on user performance or subjective attitudes. Distinct practice effects were observed. Attitude scales indicated satisfaction with the service and its interface. Finally, a significant effect of age was observed on average user response time, with the elderly activating the system quicker than the children.

ACKNOWLEDGMENTS

The author wishes to express gratitude to the many people who contributed their time and effort to the successful completion of this work. First, the members of my committee are gratefully acknowledged for their encouragement and support. The expertise, assistance, and time of Dr. Dennis Price as my committee chair is gratefully appreciated throughout all the stages of my thesis. Dr. Robert Williges provided invaluable discussion and support, as well as a useful personal library. The technical expertise, continued encouragement, and financial support of Dr. Robert Springer of NYNEX Science & Technology is appreciated.

Dr. Robert Lysaght of NYNEX Science & Technology provided invaluable advice and experience, as well as the opportunity to apply my human factors knowledge in a corporate research and development setting, expanding my personal and technical knowledge of the field.

Liz Quigley, also of NYNEX Science & Technology, is acknowledged for managing logistical issues and kindly donating her Macintosh to my experiment.

The assistance of Suzie Lee in contacting pre-school aged children was critical to the completion of my second experiment. The members of the Blacksburg chapter of the American Association of Retired Persons, the Retired Seniors Volunteer Program, and the Blacksburg community are thanked for participating in my study. Piet Kruithof is also acknowledged for his graphic design expertise and diligent work on designing emergency service icons.

Finally, I would like to thank my parents, Frederick and Josephine Kies for their support of my education and personal development.

This research is a result of support the author received from NYNEX Science & Technology and the National Institute for Occupational Safety and Health (NIOSH).

TABLE OF CONTENTS

ACKNOWLEDGMENTS iv
LIST OF TABLES vii
LIST OF FIGURES ix
INTRODUCTION 1
LITERATURE REVIEW
INTERFACE DESIGN ISSUES 4
Methods of Instruction 4
Adhesive Label 4
Template Label. 5
Icon Usage 6
Voice Prompts 7
Security 10
Method of Input Issues 11
Number of Steps 11
Use of Delimiter. 12
Distance Between Keys 13
Key Correspondence 14
LITERATURE REVIEW
EXPERIMENTAL VARIABLES 16
Issues Regarding Children 16
Issues Regarding the Elderly 17
Cognitive Loading 18
System Response Time 21
Human Response Time Considerations 23
Summary of Literature Review 25
METHODS 30
Subjects 30
Equipment 31
Questionnaire Development 32
Experimental Design 35
Procedure 37
Experiment 1 38
Experiment 2 39
RESULTS 41
Experiment 1 42
Objective Measures. 42
Subjective Measures. 50
Experiment 2 56
Objective Measures. 56
Subjective Measures. 60
Combined Data 61
DISCUSSION 68
Experiment 1 68
Objective Measures. 68
Subjective Measures. 73
Experiment 2 75
Objective Measures. 75
Subjective Measures. 76

Combined Data 77
CONCLUSIONS 80
REFERENCES 83
APPENDIXES 90
Appendix A- Percentage of Northeast states covered by 9-1-1 Systems 90
Appendix B1- Four-Step Version 91
Appendix B2- Three-Step Version 92
Appendix C- Fitt's Law Application 93
Appendix D- Information Processing Model of Emergency Service Access 94
Appendix E1- Instructional Label for Four-Step Version 95
Appendix E2- Instructional Label for Three-Step Version 96
Appendix F- Participant Reaction Questionnaire, Experiment 1 97
Appendix G- Participant Reaction Questionnaire, Experiment 2 102
Appendix H- Participant Instructions, Experiment 1 (No Cognitive Loading) 105
Appendix I- Participant Instructions, Experiment 1 (Cognitive Loading) 107
Appendix J- Participant Instructions, Experiment 2 109
Appendix K- Post Hoc Power Analysis Calculations 110
Appendix L- Questionnaire Item Analysis Plots, Experiment 1 111
Appendix M- Informed Consent, Experiment 1 114
Appendix N- Informed Consent, Experiment 1 115
VITA 116

LIST OF TABLES

Table 1. Summary of Interface Design Variables Considered 28
Table 2. Experimental Variables, Exp. 1: Elderly Subjects 37
Table 3. Experimental Variables, Exp. 2: Young Children 37
Table 4. Cognitive loading Effect on Initial Response Time Means, Exp. 1 43
Table 5. ANOVA for Initial Response Time Means, Exp. 1 43
Table 6. Cognitive Loading Effect on Average Response Time Means, Exp. 1 44
Table 7. System Response Time Effect on Average Response Time Means, Exp. 1 45
Table 8. ANOVA for Average Response Time, Exp. 1 45
Table 9. Cognitive Loading Effect on Dialing Error Rate Means, Exp. 1 46
Table 10. System Response Time Effect on Dialing Error Rate, Exp. 1 46
Table 11. Mann-Whitney U Test for Cognitive Loading Effect on Dialing Error Rate, Exp. 1 46
Table 12. Mann-Whitney U Test for System Response Time Effect on Dialing Error Rate, Exp. 1 47
Table 13. System Response Time Effect on Secondary Task Error Rate Means, Exp. 1 48
Table 14. Mann-Whitney U Test for Effect of System Response Time on Secondary Task Error Rate, Exp. 1 48
Table 15. Practice Effects Means, Exp. 1 49
Table 16. ANOVA for Practice Effects, Exp. 1 49
Table 17. Subjective Scale Items: Descriptive Statistics 51
Table 18. Item Analysis of Questionnaire Data 53
Table 19. Summary of Questionnaire Background Information 54
Table 20. Spearman Rank-Order Correlation Coefficient for Initial Response Time, Average Response Time 55
Table 21. Spearman Rank-Order Correlation Coefficient for Initial Response Time, Subscribed Phone Services 55
Table 22. Spearman Rank-Order Correlation Coefficient for Calling Errors, Subscribed Phone Services 56
Table 23. Effect of System Response Time on Average Response Time Means, Exp. 2 57
Table 24. ANOVA for Average Response Time, Exp. 2 57
Table 25. Effect of System Response Time on Dialing Error Rate Means, Exp. 2 58
Table 26. Mann-Whitney U Test for Effect of System Response Time on Dialing Errors, Exp. 2 58
Table 27. Practice Effects Means, Exp. 2 59
Table 28. ANOVA for Practice Effects, Exp. 2 59
Table 29. Subjective Scale Items: Descriptive Statistics, Exp. 2 61
Table 30. Effect of Age on Initial Response Time Means, Combined Data 62
Table 31. ANOVA for Initial Response Time, Combined Data 62
Table 32. Effect of Age on Average Response Time Means, Combined Data 64
Table 33. Effect of System Response Time on Average Response Time, Combined Data 64
Table 34. ANOVA for Average Response Time, Combined Data 64
Table 35. Effect of Age on Dialing Error Rate Means, Combined Data 66
Table 36. Effect of System Response Time on Dialing Error Rate Means, Combined Data 66

Table 37. Mann-Whitney U Test for Effects of System Response Time on Dialing Error Rate, Combined Data 66
Table 38. Mann-Whitney U Test for Effects of Age on Dialing Error Rate, Combined Data 67

LIST OF FIGURES

Figure 1. Plot of Practice Effects, Experiment 1 50
Figure 2. Plot of Practice Effects, Experiment 2 60
Figure 3. Plot of Age Effect on Initial Response Time, Combined Data 63
Figure 4. Plot of Age Effect on Average Response Time, Combined Data 65

INTRODUCTION

Telecommunications technology has advanced to the level at which a more usable and useful telephone-accessed emergency service can be offered which may have great appeal to a wide variety of potential users, including the elderly, young children, and those with disabilities. The proposed emergency service would enable a resident undergoing a traumatic experience (e.g. heart attack, devastating fire, or burglary) to activate a professionally-recorded, customized message transmitted by telephone to either the police, emergency medical services, or fire department, depending upon the situation. The caller need not remain on the line, as the message contains personalized information such as name, address, any medical problems or disabilities, easy access to the home, and where the bedrooms are located. The system serves as an alternative to traditional methods such as 9-1-1, and acts, in effect, as a "panic button" for the home.

Offering such a service, however, creates a human factors engineering dilemma. The service must be considerably easy to use when undergoing a traumatic experience, such as a heart attack or unwanted home intrusion, yet must be difficult or impossible to activate the system inadvertently . It is imperative that this human-interface predicament be resolved to ensure proper cooperation and acceptance by local emergency authorities as well as consumers.

One goal is to identify a feasible solution to the dilemma of selective usability within the constraints of a telephone key-pad. Several design configurations were examined with several segments of the population, including small children and elderly adults to gain a full understanding of potential user groups. Because the system would likely be used only in a traumatic incidence, it was important essential to assess user performance in an environment similar to an emergency situation.

It is expected that the aforementioned alternative to 9-1-1 would have considerable appeal to a wide range of consumers. One study (The Taylor Group, 1992) found that for a new telephone application, most respondents preferred telephone-accessible automated emergency services to other services such as banking, electronic yellow pages, and shopping. In advanced telephony applications, the emergency service concept was most appealing, especially for the elderly. Furthermore, the service would be a beneficial security measure for the physically and mentally disabled and desirable in the homes of small children and "latch-key" children. Various disabled segments of the population, including deaf, blind, and wheel-chair bound individuals would likely find the service useful. In fact, from a questionnaire conducted in 1991 on attitudes of the deaf toward telecommunication services, 70% of respondents rated enhanced 911 services as "highly desirable". Home security, fire protection, and automatic calls to the police were similarly rated "highly desirable" by approximately 60% of the respondents (Bowe, 1991).

It is understood that the proposed emergency service violates traditional stereotypes, namely 9-1-1. Although the increasing implementation and ubiquitous use of 9-1-1 is a trend throughout the United States today, much of the Northeast still relies on other numbers for emergency access (See Appendix A). Reversing highly-engrained habits such as the emergency telephone number has been occurring for some time now, and is thought to be feasible. However, user acceptance may be contingent partially upon willingness to adopt an alternative or supplemental method.

This research, therefore, had the following objectives: 1.) To design a feasible method to access the automated telephone emergency service; 2.) To determine the effects of system response time on user performance and acceptability; 3.) To determine the effects of increased cognitive loading on user performance and acceptability; 4.) To examine system usage by young children and the elderly.

This research was conducted in two primary phases. Over a period of several months, the interface for the service was carefully considered with special attention paid to the wealth of research previously conducted on related topics. The second phase involved the traditional experimental design and subsequent first-hand examination of human behavior.

Because the first phase involved extensive literature review closely coupled with interface design, these two aspects will be discussed together in the following section. The literature for the experimental variables is discussed in a separate section.

LITERATURE REVIEW: INTERFACE DESIGN ISSUES

Due to the potentially diverse user population and critical implications of undesirable failures-to-access and of false alarms, the method of accessing the service was carefully studied. The literature was reviewed with respect to several interface design issues. The following is a discussion of prior research that has been used to justify the decisions for excluding certain factors and including others in the system interface. The interface design factors to be discussed can be classified under three principal variables: methods of instruction, security, methods of input.

Throughout the literature review, references to memory and related terms will be discussed. In general, Wickens' (1992) model of human information processing will provide the related theoretical framework. For purposes of clarity, when short-term memory is mentioned in the literature, it will be cited as such, but equated with Wickens' "working memory". According to Wickens, working memory is defined as a temporary, attention-demanding store used to hold information until it is acted upon. Short-term memory is not to be confused with short-term sensory store, which processes information prior to working memory.

Methods of Instruction

Due to the interactive nature of telephone technology, instruction can be presented as either written instructions in the form of a label or template, voice prompts, or a combination of written instructions and voice prompts. The literature was reviewed to narrow the numerous design possibilities.

<u>Adhesive Label</u>. Large labels placed on telephone equipment by the consumer were considered to be a plausible means to transmit written instruction. One such configuration involves a large adhesive placed above the key-pad with sufficient contrast as well as the ability to attract attention. Such a label configuration is difficult to remove, preventing accidental loss, and is within the users' field of view while operating the

telephone. The label could also be placed on any convenient surface of the telephone, depending upon the style and shape of the equipment. Kramer (1968) recommends formal training for users of data entry systems requiring the use of touch-tone telephones as an input device. While this early study reflected the behavior of users not familiar with the now ubiquitous touch-tone interface, the recommendation does point to the necessity of a clear, easy-to-understand instructional scheme. One chief advantage of the large label is the capability to include relatively larger amounts of instructional information in comparison to other labeling schemes. The command card provided to users of a phone-based interface (PBI) in a study conducted by Gould and Boies (1983) met this objective by supplying examples of frequently used functions, without the unpopular heft of a larger manual. The principal drawback is that the instructions on the label are not adjacent to the keys necessary for interaction with the system. The user must read the instructions and then switch the glance down to the key-pad.

Template Label. Another written instructional scheme consists of a removable template fitting over the keys on a standard touch-tone keypad [standard dimensions provided in Woodson, Tillman, and Tillman (1992)]. The primary advantage to such a design is the proximity of instructions to the keys. For example, if the "1" were to be used to activate the police, a brief statement could correspond on the template indicating this action. Such proximity enables the user to spend less time searching for push-button keys, and relieves working-memory, as defined by Wickens (1992). One disadvantage is that limited space could prohibit detailed instructions adjacent to the keys. Furthermore, due to the variability in telephone styles available on the current market, the use of a template is restricted to those phones whose key-pad adheres to the standard dimensions. Finally, a template lacks the permanence of adhesives, and is therefore subject to loss. Prior research (Israelski, 1988) indicates that the use of an overlay containing the names of special private branch exchange (PBX) calling features in addition to an instruction card produced no significant error rate reduction.

Icon Usage. The display of information on the label raises the issue of icon usage. Literature has shown that the combination of icons with text is a highly effective method for communicating a warning message (Young and Wogalter, 1988) and instructional information (Cushman and Rosenberg, 1991). Research indicates that responses to graphical messages may be faster than speech messages (Robinson and Eberts, 1985 and 1987). Booher (1975) demonstrated that pictorial representations were helpful in increasing speed, but claimed that additional textual information was necessary for accuracy.

The combination of information presentation under different modalities has consistently shown superior performance and can be attributed to the Dual-Coding Information-Processing theory. The theory maintains that information presented in both verbal and image formats is more powerful than single representation (Paivio, 1986). According to the theory, there are two separate cognitive subsystems. One represents and processes information regarding nonverbal objects, referred to as the imagery system. The other manages language, referred to as the verbal system. Redundant modes of information presentation will take advantage of these two systems and produce a more powerful impact. Written instructions placed on an adhesive label and voice prompts (see next section) both employ the verbal system as indicated by the theory. Icons in turn make use of the imagery modality to strengthen the impact of the instructions.

Nugent (1987) supports the position that a combination of graphic and audio information is the most effective and efficient method for presenting job task instructions. When textual instructions are included, this effect is enhanced. Nugent found consistently lower performance levels for subjects whose instructions excluded the audio component. The results of this study also demonstrated that the effects of combined verbal and image codes were most profound when subjects were learning the task to be performed. In other words, once learning has occurred, the effect is no longer observed.

Voice Prompts. Another means of communicating instructional information to the user is by way of voice prompts which can be configured in a variety of ways, depending upon the resolution of other variables. It is clear that voice prompts cannot be used alone, given the variety of user hearing abilities, user acceptance, and emergency situations. Voice, however, can provide a necessary complement to instructional labels.

Applied research which supports the inclusion of a voice format is found in a study conducted by Stern (1984) in which the effects of voice messages, text, and graphics were evaluated on the use of automated-teller-machines (ATM's). Subjects who were presented with voice performed the given tasks best when the voice instructions were combined with text. Stern postulated that voice was a useful complement to text if the information is simple, and if the user population has some difficulty reading. It is interesting to note, however, that the inclusion of voice was generally disliked by the subjects in all conditions. The author noted that this may be due to the public forum of an ATM, and that if privacy is maintained, such opinions will not remain as strong. For example, it has been shown in a flight task (presumably, a non-public situation) that subjects overwhelmingly preferred the voice messages for communicating emergency procedure information as opposed to graphic and textual modalities (Hawkins, Reising, Lizza, and Beachy, 1983). One explanation is that a few, well-chosen words are less ambiguous than graphic means which may require a degree of interpretation.

A relevant point to accessing the emergency service is that information presented aurally will remain in working memory longer and is less vulnerable to interference than information presented in a visual format. This is an effect commonly found in the literature (Chapanis, 1973; Margrain, 1967; Henneman, 1952; and Savage, Kemp, and Payne, 1991). Margrain (1967) found that retention for a series of items in a list is superior when presented in an auditory mode as opposed to a visual mode. The presentation of auditory information requires time to complete the comprehension process. For this reason, Margrain argued, the auditory memory store must be more efficient over time. In a problem-solving task such as assembling a common household item or locating an address on a map, Chapanis (1973) learned that voice communication between two people was faster than any visual mode. This effect was notable even after accounting for the time to enter a visual message (e.g. typing or writing). In other words, the time to receive a spoken message was shorter than the time to read a textual message. Finally, Henneman (1952) found that when subjects performed a distracting visual or manual task, comprehension of an historical passage was increased in an auditory presentation as opposed to visual presentation. Essentially, when engaged in an attention-demanding task, aurally-presented information is more likely to be noted than visually-presented information. This is especially relevant, as mentioned earlier, because many emergency situations may require the caller to attend to the situation-at-hand or be distracted in some manner while making the call.

According to Woodson, et al. (1992), auditory information should be used when the message is short and simple, immediate action is called for, the visual system is over-loaded, and later reference is not required. Visual information, on the other hand, should be employed when the auditory system is over-burdened, the receiving environment is noisy, and information characteristics can not easily be described by words (e.g. the # sign). All of these conditions may be applicable in an emergency situation. Henneman (1952) makes the claim that auditory channels are beneficial for flexibility, selective presentation, and attention-demanding situations. Likewise, visual channels are useful for referability, a faster rate of presentation, and adaptability for presenting relational information. In general, reaction times are fastest when people are able to see and hear the stimulus simultaneously (Bailey, 1989)

Research conducted by Gould and Boies (1983,1984) and Halstead-Nussloch (1989) indicates that auditory menus in phone-based interfaces are the most appropriate means of prompting for novice users. It can be expected that nearly all usage of the emergency service will be by way of first-time or infrequent users.

Byblow (1990) makes a number of recommendations in regard to the redundancy and mode of information displayed in a cockpit environment. The claim is made that speech is an excellent method of communication when attention must be diverted to other visual cues. This is an especially relevant consideration as an emergency situation may require the user to attend to a traumatic situation while making the call. Furthermore, the presentation of non-redundant information has the potential to be hazardous because the likelihood of missing an important piece of information increases. In Byblow's study, high-redundancy levels were distinguished from low redundancy levels by the inclusion of words or pictures which revealed no new information about the state of the system. Low-redundancy versions of the graphical and textual formats were found to produce faster reaction times for both pictorial and speech warnings due to competition for scarce attentional resources. Low-redundancy levels minimized the length of the message thus avoiding working-memory overload.

Although not directly a method of instruction, various means of confirming the users' input are valid in achieving a comfortable interface in which users are able to operate the system successfully. Halstead-Nussloch (1991) and Gould and Boies (1983) reccomend employing voice confirmation messages following user actions.

Finally, the issue of speech output type can take several forms. Halstead-Nussloch (1991) recommends quality voice recordings over synthesized speech to achieve greatest legibility whenever technically feasible.

Security

Because the service allows the transmittance of personal information such as medical histories, the potential user may desire a level of security such that the information cannot be retrieved by anyone. Additionally, a security device could help prevent inadvertent actuation by young children. Unfortunately, any security measure employed would most likely hinder the speed and ease of access to the system. Unlike an ATM, the service is not expected to be used on a regular basis, disallowing the opportunity for reinforcement of a personal identification number (PIN). Such a device would likely slow, or in the worst case, completely prohibit the user from system access. It is imperative that the user be given access without a high degree of mental effort.

Research on accessing sophisticated phone services indicates that users desire the least complex method of access. Mitchell and Todd (1985) for example, compared different methods of accessing and interacting with a service, allowing the creation of phone number lists using a phone-based-interface (PBI). Subjects tended to prefer accessing the service with a minimum of coding levels. Roemer, Pendley, Stempski, and Borgstrom (1986) reccomend minimizing the number of keystrokes for functions of a voice-mail system. Finally, Gould and Boies (1983) examined a system which required users to access a voice-message system by entering their last names, rather than a meaningless code. This feature was listed as one of the "user-friendly" aspects of the system by subjects in the study. Little research directly examines the issue of security and security codes in PBI's. However, based on this research and considering the situation in which the proposed system is likely to be used, the advantages of PIN's do not outweigh the disadvantages.

Method of Input Issues

At several points throughout the access process, the user will be required to interface or provide input to the system through the telephone key-pad. Although the key-pad is considered an unattractive tool for system input (Riley, 1987), it must serve as the primary input means for the proposed service. Input opportunities involve initial system access, emergency service branch determination, and confirmation. To facilitate the literature, this input structure can be classified in terms of a number of variables. First, the number of distinct actions required for the user to send the prerecorded message must be determined. If too few are required, the chances of inadvertently activating the system may increase. If too many steps are necessary, the user may encounter difficulties. Secondly, the issue of delimiter usage must be examined. Similarly, the distance between key presses must be examined, paying close attention to the classic speed-accuracy tradeoff. Another variable of concern is that of key-correspondence. Either numeric codes or mnemonic codes could be employed as a means to identify which emergency service is to be accessed. To illustrate, the user could either enter "1" for police, or "P" for police. These issues will now be discussed.

Number of Steps. Two design configurations, each requiring a different number of steps to access the service have been considered. To activate the service, a balance must be struck between speed and accuracy in terms of the number of actions required. For example, an increase in distinct actions required may focus the user in a path (aided by voice prompts) such that access errors may be difficult to commit. However, minimizing the number of actions will reduce the time to access, an important objective.

One access configuration consisted of requiring the user to press "1" followed by the pound key to access the system initially. The user would then listen to a voice prompt with the instructions to press either 1, 2, or 3, depending upon the desired service. Thus, the user makes two discrete entries (three keys presses) to access the appropriate service.

The other configuration consisted of requiring the user to press either 1, 2, or 3, followed by the pound key, avoiding a step (see Appendix B-1 and B-2 for respective flow diagrams). Roemer, et al. (1986) and Mitchell and Todd (1985) demonstrated that subjects preferred PBI's with the least number of access steps. Additionally, multiple access tiers may cause undue stress for users already in pressured situations. The single tiered access method is hypothesized to result in no increased access errors due to the comprehensive instructional formats and the distance between the keys necessary for activation (see next section).

Use of Delimiter. When utilizing the touch-tone key-pad as an input device, the use of a delimiter key, indicating the completion of user-entered sequence to the system warrants consideration. As alluded in the previous section, the pound key (#) has been considered a possible delimiter key. Several studies indicate that PBI's requiring the use of a delimiter key are disliked by subjects (Halstead-Nussloch, 1989 and Aucella and Ehrlich, 1986). These studies subsequently reccomend avoiding such key confirmation requirements to designers of PBI's. Stuart, Desurvire, and Dews (1991) used the pound key to truncate voice prompts in PBI's and found that users neglected it, made errors, and became confused when required to use it in entering information through the telephone key-pad. If using the pound key, they reccomend explaining its use in documentation so that it can be pictured as a symbol, avoiding the confusion associated with its name.

Despite this evidence in favor of avoiding the use of the pound key in finalizing input, practical considerations warrant its inclusion in the design of the interface. First, a sole one-digit access key would not be differentiated by the network from an initial entry for a typical phone number. Secondly, the digit-pound key-sequence is the method of accessing the speed-dial features in most telephone calling regions, presenting a practical and technical advantage to implementing the service in the existing network structure. Additionally, Israelski (1989) found that users remembered a one-digit-one- symbol combination best when comparing a variety of PBX feature access schemes. Finally, when the 1-# sequence is entered, it provides a convenient method for testing the service, creating an easily programmable simulation. At the time the decision was made, the exact procedures and programming techniques were not resolved for the appropriate mock-up of the service.

Distance Between Keys. The second main issue involving key-pad input is the distance between keys. As mentioned in the preceding section, it will be necessary to require the user to enter a one-digit number followed by the pound key. The question of proximity to the pound key is of issue. The increased proximity of the bottom row of keys (7,8,9) may shave valuable time off accessing the service. However, greater proximity may increase the chance of making an error, outweighing any benefits in access speed. Alternatively, another version would require the use of the top row of keys (1,2,3). Although most touch-tone entry errors are detected immediately, of those that are not, typical mistakes include pressing keys adjacent to the correct key (Mahood, 1971). In an early study examining the problems associated with touch-tone usage, Kramer (1968) also found that one of the leading hand-keying errors was that of adjacent-digit substitution. Because the consequences of inadvertent actuation are so grave, it seems that the negligible time savings by using the bottom row would not outweigh possible keying errors. In fact, it is suggested that users would be more familiar with the location of the top three numbers (1,2,3) as opposed to the bottom three (7,8,9).

Despite this assumption, measuring the time savings between using the top row of keys and the bottom row in conjunction with the pound key offers an excellent opportunity to apply Fitt's Law. Fitt's Law originally formulated in a paper by Fitts (1954) governs the relationship between speed, accuracy, and distance. Card, Moran, and Newell (1983) used the principle to optimize the keys on a calculator, an application with evident similarities to the telephone key-pad problem. Appendix C illustrates the application of Fitt's Law to compare the estimated time difference between the top row of keys and the bottom row. Calculations predict a time savings ranging between 45.25 msec.s and 115.20 msec.s. Due to such insignificant access reduction estimates, the top row of keys is used in the final design.

Key Correspondence. The third variable of interest in designing the method of user input is that of key correspondence, of which, two versions are considered. A mnemonic version would require the user to enter a letter-assigned key for the individual service. For example, the user would enter "P" for police, "M" for medical, and "F" for fire. The alternative is to have the user enter "1" for police, "2" for medical, or "3" for fire, or a similar numeric variation. The literature offers a variety of interesting studies on the subject.

Root and Koster (1986), Reilly (1987), and Israelski (1988) found that mnemonic syntax for touch-tone entry of codes was better remembered than numeric syntax. Although Israelski did not test the mnemonic version against numeric configurations due to technical constraints, the use of mnemonics was recommended if the support for such access is feasible. Virzi (1991) also found better memory for mnemonic codes, but also noted that this was outweighed by the difficulty in locating the letters on the key-pad and the increased incidence of mistakes. Subjects expressed a strong dislike for dialing letters as opposed to numbers. Kramer (1968) also compared numeric and alphabetic touch-tone key-pad data entry under a variety of conditions and concluded that alphabetic entry should be avoided due to higher error rates. Detweiler, Schumacher, and Gattuso (1990) found that subjects had a great deal of difficulty in locating the letters on a key-pad when asked to do so from memory. Only 18% of the subjects (phone company employees) were 100% accurate in placing the letters above the correct number keys. It is likely that the general public would be even less adept and knowledgeable in this regard. Similarly, the researchers noted that much of currently available customer premise equipment have illegible lettering on the dial-pad and some exclude it altogether. Letters, however, will continue to be placed on standard telephone key-pads, as several international standards organizations have adopted standards calling for letter placement (Blanchard, Lewis, Ross, and Cataldo, 1993)

Although the evidence tends to indicate that mnemonic coding for telephone input is better remembered, research also consistently demonstrates that users have a difficult time locating the letters on a key-pad and often harbor a strong dislike for such coding. Similarly, it is expected in the case of an emergency telephone interface that users will find the search for letters on the key-pad cumbersome and time-consuming, especially under situations of high stress.

LITERATURE REVIEW: EXPERIMENTAL VARIABLES

Issues Regarding Children

The proposed emergency service is expected to play an important role in the households of a variety of users with special needs. Households with small children are included in this strategy, and the success of this service depends in part upon system mastery by young users. Therefore, one variable to be examined is that of age.

A number of special considerations must be examined when researching the usage abilities of young children. First, as Bronfenbrenner (1977) asserts, laboratory experiments have transformed developmental psychology into "the science of the strange behavior of children in strange situations with strange adults." To alleviate this problematic condition, Bronfenbrenner suggests that studies take place in the home or at school, parents or teachers act as experimenter, and the task be presented as a game or puzzle to be solved.

It is a research objective to test the youngest possible users. A recent analysis conducted by the Texas Advisory Commission on State Emergency Communications (1994) shows that children as young as 5 years old are targeted in 9-1-1 awareness campaigns. Equally interesting is that 78% of polled 5-6 year-olds were able to dial 9-1-1 correctly when presented with a touch-tone telephone. Although these figures are expected to vary among states, depending upon 9-1-1 awareness programs, the fact that the children are able to recall and dial the number indicates the potential exists to transfer that ability to the proposed system.

A cursory survey of emergency responders and related individuals over an electronic "Listserv" indicated that the age in which most young children are taught how to call emergency responders is approximately five years old.

Numerous attentional difficulties arise, however, when working with pre-school children. Shaffer (1989) maintains that when performing repetitive tasks, children less than four years old will not last more than a few minutes. At age five, children become more persistent in attempting to solve problems. Strategies, however, may be unsystematic and reflect a lack of planning. According to Miller and Weiss (1981), older children (age 13) are better than their younger cohorts at concentrating on relevant information and filtering out extraneous input that may interfere with task performance. Miller, Haynes, DeMarie-Dreblow, and Woody-Ramsey (1986) confirmed that at age six, any known problem-solving strategy may be overwhelmed by the inclination to explore other stimuli that compete for attention.

Another problem associated with young participants is the ability to read basic words, which is not realized until about six or seven years of age, but tends to vary considerably.

Finally, there is considerable variation in children's comfort levels in interacting with an unfamiliar adult. Shaffer (1989) recommends that young children be tested with a familiar adult present, preferably a parent. This familiar adult would alleviate any fears or anxieties attributable to an unfamiliar adult and situation, as well as easing any communication difficulties between the experimenter and the child.

Issues Regarding the Elderly

There have been demonstrated age-related differences in dialing time while performing complex tasks (Serafin, Wen, Paelke, and Green, 1993). Simply testing college students would likely not extract representative usability information for the system. As a group, a great deal of importance is placed on successful system operation by the elderly because they would likely constitute a large percentage of the actual users. However, as with children, the elderly offer a number of considerations which differ from common human factors research participants.

One difficulty in testing elderly users is establishing an appropriate operational definition of an "old" participant. All researchers must develop age ranges for their participants, which presents a number of problems. Primarily, age can be assessed in terms of biological age, social age, and psychological age (Birren and Cunningham, 1985) each of which can vary dramatically among participants. Despite the recognized dangers of defining a group based on chronological age in research, practical considerations often warrant such a distinction.

When obtaining data from elderly users, ethical considerations must be stringently upheld. It is especially critical that no undue psychological stress or strain be placed upon participants. Experimental settings have been shown to place additional stress and anxiety upon the elderly relative to younger participants (Eisdorfer, 1967).

Cognitive Loading

As in all thorough human factors research, it is important to understand the environment in which the system is likely to be used. As noted throughout this paper, possible scenarios under which the emergency service may be used impose numerous cognitive demands such as severely diverted attention, unusual noise and visually demanding events, and exceptional stress. Miller and Elias (1991) evaluated the use of menus for PBI's and found that user anxiety as measured by Spielberger State-Trait inventory had an important, albeit indirect link on menu performance. In general, stress has been shown to have a variety of performance impairments, among those affected are working and long term memory, decision-making, and accuracy in speeded responses (Wickens, 1992).

Short-term memory tends to be subject to a curvilinear relationship with respect to arousal (Eysenck, 1976). With moderate levels of subject stimulation, short-term retention increases. At a level of more intense arousal or stress, short-term memory performance declines. This is relevant if the user is under considerable levels of stress in a visually distracting environment and must rely on auditory cues to access the desired emergency service.

Although long-term memory is relatively unimpaired by stress, Eysenck (1976) noted that stress tends to encourage those habits which are well-learned or over-learned. This phenomenon has important significance to the emergency service in that there is the possibility that users will panic and simply dial the familiar 9-1-1. Such an occurrence is not necessarily unfavorable; however, from the perspective of a designer, it is most desirable for the product to be used when the need arises.

Studies examining the effects of emotional stress on visual pattern recognition (Simonov, Frolov, Evtushenko, and Sviridov, 1977) have indicated that increasing stress levels tend to increase pace and error rate. The trend is that highly emotional states will cause an increase in the number of perceived relevant cues in the environment. Most important, however, is the finding that high stress levels will not disallow the ability to react to an important signal.

Proper design of the system should take into account the effects of stress on frequency of errors, necessitating testing scenarios in which subjects are placed under high arousal. If, for example, it is found that high stress levels severely inhibit performance in terms of access time and frequency of errors, perhaps a more simple and well-instructed access method must be employed.

Wickens (1992) offers a number of design solutions to alleviate the deleterious consequences of stress. Training can both alleviate stress as well as reinforce the behavior such that instructions are less likely to degrade in long-term memory. Information should be reduced and well-organized. Wickens reasonably states that, "emergency procedures that must be referred to on-line must be clear and simply phrased as they will need to be followed under the very circumstances that make working memory for their contents extremely fragile." Arbitrary symbolic coding is to be avoided and proceduralized instructions should be redundantly coded.

In regard to the elderly, Kausler (1991) reviewed the literature on the effects of arousal and indicated that many researchers tend to disagree. For example, Botwinick and Kornetsky (1960) and Thompson and Nowlin (1973) maintain that elderly subjects can actually become under-aroused when stressed. Others, however, maintain that arousal measures may increase more for the elderly or remain the same as younger subjects under stress (Backman and Molander, 1986a and 1986b). To explain these differences, it is argued that because autonomic arousal involves a number of physiological components, some of these may be over-aroused in the elderly, while others may be under-aroused (Woodruff, 1985). Backman and Molander showed how arousal as measured by heart rate and a subjective rating hindered performance for the elderly more so than younger subjects on a precision sport task while heart rate actually increased at the same rate for the two age groups. Additionally, lower skill levels in concert with the aging process are penalizing when precise motor-action is required under high-stress conditions.

It is recognized that simulating a high-fidelity environment in which stress is placed upon the participant is impractical as well as unethical. Additionally, relative to younger subjects, elderly subjects have become over-aroused in experimental settings (Eisdorfer, 1967).

In reviewing various "stressor" tasks, Spieth (1965) identified three levels of "pacing stress." In the least stressful, self-paced task, subjects are told to perform rapidly on a task that stands still, or does not progress as quickly as the subject may desire. The step-paced task requires the subject to respond within a certain time limit. Wrong or omitted responses have no impact upon the remainder of the task. The third, and most stressful task is the super-paced task. In this task, the subject performs a tracking task which when one performance segment is incorrect, the next segment becomes more difficult, resulting in an "out-of-control" situation.

Spieth (1965) maintains that those participants subject to "emotional disorganization" are least penalized by the self-paced task, and perform the worst on the super-paced task. Subjects of both high and low anxiety levels demonstrated practically equal performance for the step-paced task. Because there is a recognized interaction between psychometrically-determined anxiety levels and the type of task (self-paced versus super-paced), the step-paced task was used for the reason that subject anxiety level will not be controlled.

Much of the dual-task literature points to the phenomenon that task combinations involving well-rehearsed or automatic routines tend to result in smaller interference effects as compared to tasks involving controlled processing (Kortelling, 1994 and Schneider and Fisk, 1982). Cerella (1982) reviewed a large number of published dual-task studies and concluded that the slowing of sensory-motor processes is less pronounced than the slowing of higher-order processes. Such a conclusion is relevant to the choice of secondary task . Salthouse (1982) explains the phenomenon of slower behavior in the elderly by claiming that a dual task is, in effect, a single complex task which requires more mental operations. Such a slowing effect is caused by the increased complexity, and labeled the "Slowing-Complexity Hypothesis."

System Response Time

Currently, it has not been determined if "Rapid Connect" technology can be employed in the system enabling nearly instant access to voice prompts or confirmation messages. If such technology is not available, system response times as long as ten seconds could be introduced between the final key press and the first voice connection. System engineers would benefit from user data indicating tolerance for a delay at this interval and its potential impact on task performance. Little behavioral research exists directly indicating the effects of such a delay on human-telephone interaction.

Shneiderman (1987) discusses three chief factors influencing response time acceptability. First, prior experiences are critical to how users perceive a system. If a system completes a task faster than a user expects, the user is pleased. If, on the other hand, the system takes longer than the prior experiences of the user, negative attitudes will result. Secondly, Shneiderman makes the point that users' tolerance for delays varies considerably, depending greatly upon the task and the individual. It is difficult to extend an optimal response time to a wide range of tasks. Additionally, novice users may be more tolerant to longer delays than expert users. Finally, Shneiderman makes the dubious claim that users are adaptive and can change working style and strategy to accommodate longer response times (which may impact performance due to limitations such as memory capacity).

The human-computer interaction literature is somewhat mixed in regard to recommending delay tolerances and predicting effects on task performance. A few studies suggest that time delays generally longer than two seconds are unacceptable (Miller, 1968 and Williges and Williges, 1982). Miller makes user-tolerance recommendations for various system functions. For example, initial dial-tone response time may lag as much as three seconds. A routine computer request should take no more than two seconds. Williges (1983) recommends that optimal systems should acknowledge a request nearly instantaneously (e.g., <.5 seconds). Based on these studies, it is suggested that the emergency system users would have limited patience for time lags longer than two seconds.

Other studies, however, indicate that longer delays are acceptable and tolerable. Murray and Abrahamson (1983), for example, showed that for videotex tasks, constant delays as long as 10 seconds result in no significantly different attitudes by novice users. Delay variability plays a much larger role in influencing users' subjective attitudes toward such a system.

While Bergman, Brinkman, and Koelega (1981) suggest that much work has been done on subjective reactions toward system response times, little research has examined objective performance measures. Bergman, et al. researched the effects of system response time on a complex problem solving task. It was concluded that system response time had no effect on task performance and that the specific task is the most important determinant of delays' effect on performance.

Weiss, Boogs, Shodja, and Martin (1982) examined the effects of system response time on a process control task. Delay was shown to have no effect on physiological measures such as heart rate and blood pressure. However, longer system response times resulted in decreased error rate. Weiss, et al. further reccomend that error rate be a key dependent variable in future research.

Butler (1983) also measured the effects of system response time on different tasks (a simple and complex data entry task). Butler found no significant effect on error rate, but noted that user response time did tend to increase with marginal statistical significance as the system response time increased. This finding has important implications for this study because user response time is one of the primary dependent measures, as will be discussed in a later section.

Human Response Time Considerations

Wickens' model of human information processing (1992) will be applied to the behavior of subjects in the study. Human response time will be differentiated from reaction time by including all the information processing events leading up to and including response execution. In typical studies referenced by Wickens, reaction time is terminated at the initiation of the task (e.g. pressing a button). In the case of this study, the time was measured to the completion of the task (pressing the pound key). The rationale for making this distinction rests in the salience of measuring the time to complete the task in a realistic situation, as well as technical constraints associated with the experimental prototype.

The major components of Wicken's (1992) model of information processing which characterize the access process include sensory processing, perception, decision and response selection, and response execution. Feeding into these components are other cognitive elements such as attentional resources and the various memory stores.

The task considered in this paper also requires an added degree of complexity in that two actions are required to cause the service to be successfully activated: as mentioned in earlier sections, the appropriate emergency branch must be selected with the correct key press, followed by pressing the pound key. Response time, in this context, therefore includes the time taken for all the information processing events, as well as the time to press both keys in the activation sequence. Appendix D provides an adaptation of Wicken's model to the task of accessing the emergency service.

In deference to the above operational definition of response time, reaction time warrants discussion for purposes of clarification. Many factors affect one's performance on reaction time tasks (Wickens, 1992); however, a few are especially relevant in the context of an emergency situation.

Woodsworth and Schlossberg (1954) make the case that stimulus modality impacts reaction time. For example, auditory stimuli result in reaction times approximately 40 milliseconds faster than visual stimuli. By contrast, some researchers (Kohfeld, 1971) claim that prior research did not control for stimulus intensities among varying modalities. Kohfeld compared psycho-physically equivalent auditory and visual signals and found that at higher-intensity levels, there was no reaction-time difference, but at low levels, the auditory stimuli produced shorter reaction times, consistent with earlier findings.

Because the results in this area tend to be somewhat controversial, the stimulus presentation was consistent across all conditions and subjects. Subjects were exposed to a recorded verbal message indicating which emergency service is required. This also prevented differences in reading ability and speed among subjects from influencing the dependent measures.

A second factor impacting reaction time relates to the classic speed-accuracy tradeoff (Wickens, 1992) in which the manipulation of instructions to the subject have a proven effect on reaction time. Strayer, Wickens, and Braune (1987) found that when older subjects were instructed to maximize speed, the increase in errors was significantly fewer than the increase in errors by the young subjects. Furthermore, when instructed for accuracy, there was little difference in number of errors among the different age groups. Overall, however, there was a 150 ms. increase in reaction time between the youngest and oldest groups. For these reasons, the issue of speed-accuracy instructions must be carefully considered.

The final relevant issue affecting reaction time is that of repetition. It is not surprising that a commonly reported finding in the literature (Wickens, 1992) is that with repeated trials, reaction times will shorten. In other words, repeating a stimulus-response pair will result in faster times than using an alternation. This finding is important as several trials were taken for each subject, but it is likely that only the first trial provided the most valid data in that a realistic situation would not allow a user to access the system repeatedly in a short period of time.

Summary of Literature Review

Prior research has been used to make design decisions for the emergency service as well as define which variables require further experimental investigation. The following is a summary of the variables considered coupled with a brief rationale for their inclusion or exclusion from the experimental design.

After weighing the relative advantages and disadvantages of the two labeling schemes (adhesive and template), it was decided to employ the adhesive version with some aspects of the template. A large label located above the key-pad contains basic activation instructions, while smaller labels signify the individual keys (see Appendix E1 and E2 for illustration). The research and clear practical issues associated with each labeling scheme indicate that testing this variable would not make efficient use of experimental resources.

According to the research, the best design for label layout is to include icons or pictures in concert with textual and/or vocal means of information presentation. Furthermore, the stressful nature of many emergency situations warrants the instructions maintain a high degree of clarity, necessitating the use of written instructions. Icons, however, can be helpful in terms of further clarification by indicating police, fire, and medical with commonly-used and accepted pictograms. Likewise, many customer premise equipment (CPE) models with speed-dial features currently make use of these icons. For these reasons, the layout of the label was not varied in the experimental design. Additionally, natural voice output was used rather than synthesized voice for purposes of clarity as indicated in the literature.

Given the research, voice was included as an instrument for confirmation, rather than a means of instruction. As noted earlier, redundant information is especially useful in a situation in which attention may be diverted to other tasks. When a user accesses the service, it is likely that a fire, medical, or security problem is occurring which would place demands on the attention of the caller. Furthermore, the service is not likely to be accessed on a regular basis, thus it can be argued that virtually every usage can be regarded as a first-time experience, necessitating simple, clear, and effective communication for the methods-to-access. Finally, the user population may be heavily weighted by individuals with a variety of disorders and disabilities which may further be hampered by the stress of an emergency situation. For these reasons, voice prompts were included as a feedback tool in the emergency service design. It was not, however, used as an instruction tool, because it would likely frustrate the user and make the activation process more complex than necessary.

The use of security devices was excluded from the design due to the obstacles they impose upon successful and timely access. A questionnaire item, however, addresses this issue.

With respect to methods of input to the system, requiring four steps to access, as well as a vocal instruction message was deemed too cumbersome. A variety of research indicates that users prefer the least number of tiers in an access scheme. For these reasons, the single-digit, pound sequence formed the basis for accessing the system.

The delimiter key was included in the access method, resulting in the requirement of a number-pound sequence. Despite mixed research on users' opinions of delimiters, network architectures dictate the use of such a key in the described system.

In regard to the key-distance variable, the top row of keys will be used for determining which service is desired due to their relative isolation from the other keys, reducing the chance for inadvertent key substitution. Furthermore, the increased distance from the pound key would likely not increase the time to activate by an appreciable margin, as determined by a Fitt's Law analysis.

The key-correspondence issue discussed earlier was not examined in the experimental design. Due to the substantial evidence pointing to difficulties locating letters on a telephone key-pad, the numeric key-correspondence version will be applied to the emergency service design. Table 1 provides a summary of the interface design variables considered to this point.

TABLE 1

Summary of Interface Design Variables Considered

Methods of Instruction	Security	Methods of Input
Adhesive Label		Number of Steps
Template Label		Use of Delimiter
Icon Usage		Distance between Keys
Voice Prompts		Key Correspondence

Several non-design issues were discussed in the literature review, affecting the experimental conditions under which the system prototype was tested.

First, the abilities and testing considerations involving extreme age groups were examined. It was determined that the youngest useful age group would be in the 5-6 age range, pre-kindergarten. In researching the elderly, a great deal of variability resides in determining a person's age (e.g. biological, psychological, etc.), but for purposes of this experiment, the elderly are defined as older than 60 years of age.

The inclusion of an environment in which attention is diverted from the primary task of making the call will be included as an independent variable to ascertain an indication of how the system may be used under realistic, cognitively-demanding situations. Literature has shown that various levels of stress can influence such relevant constructs as working- and long-term memory and reaction time. The elderly are expected to perform two tasks simultaneously, but pilot testing revealed that a secondary task cognitively overwhelmed young children. Thus, the cognitive loading variable was only presented to older participants.

The nature of network switching systems and architectures is such that time delays in connecting the user with voice prompts may vary considerably. The literature regarding the effects of system response time on user performance and subjective attitudes was reviewed. Due to inconclusive findings in the literature and the need to aid system designers in understanding user-acceptance for this domain, an independent variable of system response time was experimentally manipulated.

Because the system requires a user to process information and react by pressing keys on the key-pad, the issue of user response time is relevant. It was concluded that all stimuli prompting the participant to react and call one of the emergency service branches be presented in the same modality. For purposes of consistency, that modality will be human voice-generated.

Finally, because considerable weight is being placed upon subjective responses as a means for assessing the different emergency service versions, related research was reviewed (see Methods section) to enable the successful creation of a useful questionnaire. A separate questionnaire was constructed for young children to account for their different language and cognitive abilities.

METHODS

Subjects

To test the different design configurations adequately and ensure valid results, the subject pool was divided into two distinct age groups. In all cases, it is assumed that small children under the age of four would not be able to activate the system inadvertently. The age groups were separated into the following categories:

5-6 Years Old. This age group was tested to generalize the youngest population of potential users. Sixteen subjects were tested with a mean age of five and a range of four to six. Children were compensated for their (parents') time and effort with a Virginia Tech T-Shirt.

For this research, pilot testing with small children indicated that the instructional label was essentially ignored. However, the individual labels located above each service branch key were noticed, although perhaps not read. The first letter of each word was printed larger and in bold typeface to capitalize on the tendency of young children to utilize "phonics" or individual letter sounds to recognize the correct emergency service.

Originally, cognitive loading was proposed as an independent variable in the study to gain an indication of how the system might be activated under more realistic circumstances. Pilot testing, however, revealed that children at five years of age have considerable trouble paying attention to more than one task. When instructed to perform the simple secondary task in concert with the main task of making the calls to the emergency services, the secondary task was essentially ignored. For these reasons, this condition was eliminated from the study of young children.

Elderly. It is acknowledged that this segment of the population should be highly comfortable with the operation of the system and have no trouble in its activation. Relatively independent, self-sufficient, and healthy elderly subjects were chosen from members of the Blacksburg Chapter of the American Association of Retired Persons

(AARP), the Retired Senior Volunteer Program, and others in the community who responded to a newspaper advertisement. Thirty-two subjects were tested with a mean age of 70 and a range of sixty to eighty-one. Adults were compensated with their choice of five dollars or a Virginia Tech coffee mug.

It is fully acknowledged that such a subject pool may not constitute a representative sample of potential elderly users in terms of health, attitude, vigor, intelligence, and economic background. Additionally, such participants may reflect characteristics of much younger users, as evidenced by willingness to participate in the study.

Equipment

The experimental stimulus involves a prototype of the emergency service, which participants dialed into over standard phone lines. Subjects were presented with a working telephone of the standard touch-tone, "no-frills" variety. An adhesive label located above the key pad included the instructions for operation as noted earlier.

Prior to the subjects' trials, the telephone was pre-dialed and connected to the line in the Safety and Environmental Laboratory on the 5th floor of Whittemore Hall on the campus of Virginia Tech. The line was linked to the TFLX unit (manufactured by Magnum Software Crop.) and a Macintosh computer (Apple Computer Inc.). TFLX is a telephony-oriented application development tool which was configured such that the emergency service could be prototyped and accessed. In the conditions for which a secondary task was to be administered (cognitive loading conditions), a second Macintosh running a HyperCard program was used. Finally, a questionnaire was presented to the subjects at the end of the session (Appendix F and G).

When running subjects at Virginia Tech, the usability room in the Environmental and Safety Laboratory was utilized. The room offers minimal distractions for the session.

Some sessions were also conducted at the Blacksburg Baptist Church in a partitioned corner of a large hall.

Questionnaire Development

Due to the weight placed on the subjective measures as means for assessing the effectiveness and overall performance of the two emergency telephone versions, a few comments on the questionnaire development are appropriate.

The first three sections of the questionnaire consist of simple background and product familiarity questions used to gain an understanding of the subject's experience with relevant technology and to help categorize subjects' performance by demographic factors (see Appendix F). The fourth section of the questionnaire consists of a series of questions which will be used to assess user attitudes and opinions toward the system.

These questions are modified from those included in the Questionnaire for User Interface Satisfaction (QUIS 5.0) as presented by Chin, Diehl, and Norman (1988) and Harper and Norman (1993). QUIS has a reported reliability of .94 as measured by Cronbach's Alpha, has high internal consistency, and is reported to be externally valid, although these statistics cannot be extended to the questionnaire used in this study. The QUIS items are arranged as 10-point scales combining features of Likert scales as summarized by DeVellis (1991) as well as the semantic differential scaling method as described by Osgood, Suci, and Tannenbaum (1957).

The primary attribute of QUIS, which was borrowed from traditional Likert scales, is that the participant responds to a statement by selecting a number on a continuum, expressing a reaction towards that statement. Rather than measuring agreement or disagreement with a declarative statement, QUIS measures various attitudes toward a characteristic of the system. The questionnaire used in this study makes use of this questioning style.

Hix and Hartson (1993) suggest that many developers supplement QUIS with other questions relevant to the interface being assessed. The questionnaire used in this study has been modified from QUIS to incorporate some of the original characteristics of semantic differential scaling as outlined by Osgood, et al. (1957), Nunnally (1967), and DeVellis (1991) and include questions more appropriate to the emergency telephone service.

Osgood, et al. (1957) used a factor analytic technique to cluster bipolar adjectives into common factors. Although numerous factor analyses have been conducted on semantic differential scales with varying results, three prominent factors commonly surface, according to Nunnally (1967). Osgood, et al. (1957) described *evaluation* as the factor accounting for the majority of common variance among the bipolar adjective clusters. *Potency* and *activity* are the other two commonly cited factors in the literature. These three factors have shown to generalize semantic space robustly across cultures, as summarized by Osgood (1962). According to Nunnally, the evaluative factor nearly acts as a definition for the term, "attitude", making scales weighted in this realm appropriate to the measure of attitudes. The items included in the proposed questionnaire strive to capture the evaluative nature of users' attitudes.

QUIS makes use of a discrete numeric scale anchored with bipolar adjective pairs, requiring the user to circle the number which most closely corresponds to his or her attitude to a given statement. Due to the discrete nature of the scale, however, parametric statistical tests are not appropriate. Madden and Bourdon (1964) manipulated rating scales on a number of factors, including the inclusion of a graphic "bar". Although the researchers reported statistically significant effects of this feature, the differences were too small to make vigorous recommendations. Blumberg, De Soto, and Kuethe (1966) suggest that graphic or numeric versions of rating scales have no impact on responses. For these reasons, the proposed scales will include a straight line underneath the numbers, attempting to preserve the continuous element, while not greatly impacting the nature of the responses.

Each QUIS scale and this study's scale items are presented such that the left-most value is anchored by a negative word, and the far right is anchored by a positive word or phrase, although Blumberg, et al. (1966) show this makes no difference. The scale items included in this study's questionnaire have been amended from Chin, et al. (1988) to make use of anchor words which are literal opposites (e.g. satisfactory vs. unsatisfactory; pleasing vs. unpleasing) as recommended by Meister (1985). Retaining literal opposite anchor points ensures that each individual scale measures only one underlying variable, avoiding validity problems associated with multiple dimensions represented in one scale. In cases for which literal opposites were not grammatically correct or sensical, items were borrowed from other studies involving questionnaire development (Coleman and Williges, 1985; Bailey and Pearson, 1983).

In regard to the number of levels per scale, Finn (1972) suggests that subjects are unable to discriminate more than seven levels. Jenkins and Tabor (1977) make the case that scales with more than five levels do not increase reliability significantly. Providing fewer than five levels, however, creates a measurement tool not fine enough to detect potential discriminations, as well as making statistical analysis problematic. This study's scales include seven levels of discriminability, as suggested by the literature.

Although the need for validity and reliability in a scale are evident, the steps necessary to ensure such qualities are arduous and time-consuming, and therefore have not been taken for the proposed scales.

Finally, to ensure that participants use the scales unambiguously, enabling simpler analysis, an example was given and instructions were provided to circle the number on the scale which corresponds most closely with their attitude.

When testing young children on the system, a different questionnaire was constructed to take into account the cognitive abilities indicative of their age. Pilot testing demonstrated that nearly all the items on the questionnaire in Appendix F were not comprehensible by young children, as expected. Some items were simply deemed inappropriate. Pilot testing also revealed that children can only discriminate best on three point scales. For example, the following question: "How did you like the telephone? would have three possible response choices (Did not like, did not care/know, and liked a lot). See Appendix G for the complete children's questionnaire.

Experimental Design

Originally, the proposed study intended to expose the elderly and young children to two levels each of two factors, producing the following between-subjects factorial design: System Response Time (2) x Cognitive loading (2) x Age (2). Following pilot testing, however, it was concluded that the secondary task was more than most young children could process effectively. For this reason, the proposed study has been subdivided into two separate data-collection experiments with similar controls. The first experiment involved the effects of system response time and cognitive loading on elderly users, producing a 2 (cognitive loading) x2 (system response time) between-subjects design with eight subjects per condition. The second experiment entailed exposing young children only to the two levels of system response time, again with eight subjects per condition. The dependent variables for both routines included time to access the system, number of errors committed, and questionnaire responses (See Table 2 for summary).

Because subjects in these age groups are difficult to obtain on a college campus and the testing different testing facilities were utilized, subjects were not randomly assigned to conditions. Rather, the convenience of having the proper computer equipment on site dictated which subjects were placed in the conditions. It should be noted, however, that no systematic assignment was made on the basis of demographics or other personal characteristics of the participants in the study.

System response time is defined as the amount of time between pressing the last button in the dialing sequence and the point at which the confirmation voice message is heard by the user. The first level of this variable was set at two seconds, the tolerance threshold for computer feedback as indicated in the literature review. The other level was set at eight seconds, the longest expected system delay. The cognitive loading variable has two levels: absence and presence of the secondary task (see procedure section for full explanation).

In regard to dependent measures, the time to access the system was operationally defined as the time required for the user to respond and react after listening to the recorded scenario. The interval between the end of the voice prompt and the final action of pressing the pound key (#) was measured as "time to access the system". This interval was measured automatically and subsequently imported to an electronic spreadsheet by the software driving the telephone service . Additionally, the average response time was calculated for each subject by summing the response time for each of the six trials and dividing by the number of trials. In measuring the number of errors, an error was defined as the number of mistaken key presses (not counting those required to reorient the user on the system). These operationally defined measures will apply to all trials for both experiments. Tables 2 and 3 display the independent and dependent variables for experiments one and two, respectively.

TABLE 2

Experimental Variables, Experiment 1: Elderly Subjects

Independent Variables (Between-Subjects)

System response time --- (2 seconds, 8 seconds)

Cognitive loading --- (absence, presence of secondary task)

Dependent Variables

Response Time

First trial

Average across first six trials

Number of Errors (Number of incorrect keys pressed)

Number of Errors on Secondary Task (Cognitive loading condition only)

Questionnaire Responses

*8 subjects per condition = 32 subjects total

TABLE 3

Experimental Variables, Experiment 2: Young Children

Independent Variable (Between-Subjects)

System response time --- (2 seconds, 8 seconds)

Dependent Variables

Response Time

First trial

Average across first six trials

Number of Errors (Number of incorrect keys pressed)

Questionnaire Responses

*8 subjects per condition = 16 subjects total

Procedure

Prior to collecting data, brief, informal pilot testing was performed to ensure that the experimental procedure would take place in a well-executed fashion. For the first experiment, two elderly subjects and two college-aged subjects walked through the experimental protocol, allowing last-minute adjustments to be made. In the second experiment, three 5-year old subjects were exposed to the apparatus, similarly permitting refinements and providing a needed understanding of children's capabilities at this age.

Experiment 1. Following a brief verbal description of the study and the reading and signing of informed consent forms, an instruction sheet written in simple language (see Appendix H) was read aloud to the subject while he or she read silently. The instruction sheet provided a description of the system, its uses, and most importantly, the method for accessing the service. Any questions were addressed at that time. A touch-tone telephone was presented for those who were not readily familiar with this style of telephone. All subjects were asked to wear corrective glasses, if necessary.

When the subject indicated that he or she was ready to begin, the cycle of trials was initiated by the experimenter, who remained in the room throughout the session. The cycle adhered to the following format: A voice prompt was broadcast over the speaker phone instructing the subject to call a specific emergency service branch. The subject lifted the handset, called the appropriate branch, and after a given time lag, a confirmation message was played. Following the confirmation message, the subject was instructed to hang up the phone. After a few seconds, the next voice prompt instructing the subject to call another emergency service was broadcast over the speaker phone. Following six trials, the session paused while the experimenter explained that the seventh trial would result in a different voice confirmation message (a more realistic message was played) Following the seventh trial, the experimenter explained the use of the stop feature (* key) and instructed the subject to cease the eighth and final trial following activation. This trial was used to gain an understanding of how easily the participant could cease activation of the system. After all trials were completed, the subject was presented with the questionnaire.

The conditions in which subjects were exposed to the secondary task followed a slightly different protocol. After being read nearly the same instructions (see Appendix I) and all questions had been addressed, the subject was seated in front of the Macintosh running the HyperCard cognitive loading program. The subject was instructed to count the number of vertical bars on the screen by holding up the number of fingers for each bar. The instructions also indicated that performance on both tasks is expected to be maximized, but the priority should lie in making no errors dialing the phone. The screen refreshed every three seconds with an apparently randomly generated number of bars between one and four. A one second gap between the display of the bars was used to help indicate that a new display is to occur. Brief pilot testing of two (different) representatively aged subjects showed that a shorter refresh rate of one or two seconds became overly confusing and taxing. In the interest of minimizing the stress imposed upon the elderly subjects, the slower refresh rate of three seconds was chosen. The sequence was originally generated in a random fashion, but was repeated for each subject. Predetermined screens enabled the experimenter to score performance on the task easily. After ten cycles of the program, the voice prompts broadcast over the speaker phone were initiated. The remainder of the experiment adhered to the outline for the non-secondary-task-condition, as provided above. Finally, any questions were answered and the subject was paid and excused.

Experiment 2. In the second experiment, the subject pool consisted of young children being exposed to one of two system response time levels of the telephone system, with no cognitive loading condition. The experimental procedure followed that of the first experiment with a few exceptions.

Each child was separately taken into the experimental room with a parent. Because parental consent was required prior to testing, the child was simply instructed that he or she may quit the experiment at any time. The instruction sheet in Appendix J was read aloud and any questions were addressed. A touch-tone telephone was presented for those who were not readily familiar with the style of telephone. During the instructions, the activation process was briefly demonstrated, unlike the instructions for the elderly. The experimenter lifted the handset, and pretended to press the buttons on the phone, but not actually activate the service. Pilot testing proved that children of this age have considerable trouble reading and solving this novel problem without prior explanation, and must be explicitly instructed in the use of the phone. The child was not, however, given the opportunity to dial or activate the system at this point.

Once any questions had been addressed and the subject indicated readiness, the experimenter initiated the cycle of voice prompts. The voice prompt cycle followed the first experiment exactly. After seven trials, the subject was told how to stop the system in the event of a mistake. The eighth trial was initiated to allow the subject the opportunity to cease the system.

Once all trials had been completed, the child was presented with the questionnaire provided in Appendix G. The experimenter read the questions aloud while the scales were shown to the subject. Prior to dismissal, the subject was told how to call 9-1-1 at home and that the emergency service does not work any where else. Any final questions were addressed, the subject was paid, and dismissed.

RESULTS

The results from the two experiments will first be presented separately and then in a combined manner. Both subjective and objective measures will be presented. In all analyses, an alpha level has been set at .05 for reasons of convention and a lack of grave consequences associated with making a Type I error. It is acknowledged that establishing such a "go/no-go" distinction is fundamentally arbitrary, and therefore, non-significant p-values approaching .05 will also be analyzed with a post hoc Power Analysis to provide the reader with an indication of how many additional subjects would have been required to reach a significant difference between treatment means.

Prior to presenting the results of the study, a discussion on the use of parametric and non-parametric statistical procedures is warranted. Assumptions made in conducting tests such as Analysis of Variance are that the sampling distribution is approximately normal in shape, there is homogeneity of variance, and scores were independently measured (Keppel, 1991). Cochran (1971) and more recently, Keppel (1991) reviewed numerous studies to evaluate the problematic conditions resulting from violations of these assumptions. Both researchers maintained that reasonable violations did not have profound impacts on statistical significance, especially when using the relatively robust analysis of variance. Nonetheless, the distributions in this study were examined and checked for violations of these assumptions.

In all cases, it can be inferred that the third assumption was met, as the data collection procedures outlined in the previous section indicate no biases or systematic confounding of variables. To assess the normality of the distributions sampled, the Kolmogorov-Smirnov one sample test was employed as a goodness-of-fit test to compare the observed distribution with the theoretical (predicted) normal distribution. Finally, the assumption of homogeneity of variance was assessed by calculating the F_{max} statistic, comparing the difference between variances of the samples under consideration. Those distributions not exhibiting homogeneity of variance or approximating the normal distribution were subsequently analyzed with distribution-free tests.

Experiment 1

Objective Measures. Four primary objective measures were used to help determine the subjects' performance in regard to the emergency service. Response time was measured two ways. The initial response time (e.g. the first trial) was treated as one measure, and the mean of the response times over the six trials was treated as another measure.

The third general objective measure was the number of errors committed while dialing the phone. The final objective measure was the number of errors committed on the secondary task , to which only half the subjects were exposed.

An important note about the measurement of the response times should be made. The software driving the emergency service prototype measured the response time as an interval between the end of the voice prompt and the point at which the pound key was pressed in the dialing sequence. Thus, the measure of response time is somewhat of misnomer in that it includes the time taken to press both keys, not just the first key in the sequence. Secondly, a degree of noise was introduced in the measurement of the interval as a function of the voice prompt recording. Little can be done to cut the recording at literally the exact moment the utterance is complete. Care was taken to keep this additional time negligible, but its presence was difficult to eliminate entirely. Finally, when responding to a voice prompt, such as "Call the Police Department," many subjects knew which key to press before the statement was completed. For example, in the above prompt, following the mention of "police," the subject would know to call the police and may initiate actions toward such a request.

Table 4 provides the descriptive statistics for the effect of cognitive loading on initial response time. The system response time variable is not displayed as no significant differences could be expected for the initial response time (the system delay is incurred following the initial user response).

Table 5 provides the analysis of variance summary table for the two independent variables in regard to the dependent measure of initial response time. As the table indicates, the effects of cognitive loading and system response time (demarked in all tables as "SRT") did not significantly impact the initial response time as p-values are considerably higher than .05 or .10.

TABLE 4

Cognitive Loading Effect on Initial Response Time Means, Experiment 1

	Count	Mean	Std. Dev.	Std. Error
No Cog. Load	16	3.682	1.843	.461
Cog. Load	16	4.768	3.025	.756

TABLE 5

ANOVA for Initial Response Time, Experiment 1

Source	df	Sum of Squares	Mean Square	F-Value	P-Value
Cog. Load	1	9.440	9.440	1.436	.2408
SRT	1	.149	.149	.023	.8816
Cog. Load * SRT	1	4.061	4.061	.618	.4384
Residual	28	184.007	6.572		

Dependent: Initial Time

SRT = System Response Time

All ANOVA values for which a p-value less than .30 but greater than .05 will be re-evaluated by way of a post hoc Power Analysis, as described by Keppel (1991). The purpose of this analysis is to provide an indication of how many subjects would have had to be tested in order to achieve statistical significance. Only those F-values within a 5.01% -30% chance of reflecting a Type I Error (incorrectly failing to reject the null hypothesis) will be evaluated due to the unreasonably large sample sizes required by differences with p-values larger than .30 for a sensible level of power (as evidenced by the Pearson-Hartley charts cited in Keppel). In each case, a power level of .80 is sought, as this is considered to be a "reasonable" level of power and raising power towards the .90 range requires a considerable increase in sample size. Alpha levels will be maintained at .05 for all power analysis calculations.

As indicated by Table 5, the effect of cognitive loading on initial response time was close to statistical significance F (2, 28) = 1.436, p = .24. A post hoc power analysis reveals that nearly 155 subjects would had to have been tested to achieve a significant difference with alpha = .05, power = .80. See Appendix K for formulas and details of the calculation.

Tables 6 and 7 display the descriptive statistics for the effects of cognitive loading and system response time on average user response time. Table 8 presents the 2x2 between-subjects analysis of variance values for the two independent variables with respect to average response time. Clearly, the table reflects a failure to reject the null hypothesis.

TABLE 6

Cognitive Loading Effect on Average Response Time Means, Experiment 1

	Count	Mean	Std. Dev.	Std. Error
No Cog. Load	16	2.668	.825	.206
Cog. Load	16	2.897	1.328	.332

TABLE 7

System Response Time Effect on Average Response Time Means, Experiment 1

	Count	Mean	Std. Dev.	Std. Error
Two Seconds	16	2.609	.775	.194
Eight Seconds	16	2.955	1.345	.336

TABLE 8

ANOVA for Average Response Time, Experiment 1

Source	df	Sum of Squares	Mean Square	F-Value	P-Value
Cog. Load	1	.421	.421	.332	.5692
SRT	1	.956	.956	.753	.3928
Cog. Load * SRT	1	.194	.194	.153	.6989
Residual	28	35.527	1.269		

Dependent: Average Time
SRT = System Response Time

Tables 9 and 10 display the descriptive statistics for the effects of cognitive loading and system response time on dialing error rate. Table 11 provides analysis of variance data for the independent variables with respect to dialing error rate. Dialing errors were considered the number of incorrect key presses which coincided with one of the following scenarios: 1) Incorrect emergency service branch called (regardless of correction) or 2) Omission of the pound key in the dialing sequence. Each subject called the service 6 times, with two key presses per activation, creating a total of 12 key presses. Only three subjects committed one error each in the cognitive loading condition compared with no errors in the non-cognitive-loading condition. Informal examination of the distribution of calling errors, as well as results from the Kolmogorov-Smirnov One-Sample Test to detect deviation from the normal distribution indicate that considerable violations exist, warranting the use of a non-parametric statistical analysis. Table 11 and 12 provide the non-parametric Mann-Whitney U analysis data for the effects of Cognitive Loading and System Response Time on Dialing Error Rate. The p-values suggest that no significant differences resulted from the manipulation of the two independent variables.

TABLE 9

Cognitive Loading Effect on Dialing Error Rate Means, Experiment 1

	Count	Mean	Std. Dev.	Std. Error
No Cog. Load	16	0.000	0.000	0.000
Cog. Load	16	.016	.034	.008

TABLE 10

System Response Time Effect on Dialing Error Rate Means, Experiment 1

	Count	Mean	Std. Dev.	Std. Error
Two Seconds	16	.010	.028	.007
Eight Seconds	16	.005	.021	.005

TABLE 11

Mann-Whitney U Test for Cognitive Loading Effect on Dialing Error Rate, Experiment 1

U	104.000
U Prime	152.000
Z-Value	-.905
P-Value	.3657
Tied Z-Value	-1.791
Tied P-Value	.0733
# Ties	2

TABLE 12

Mann-Whitney U Test for System Response Time Effect on Dialing Error Rate, Experiment 1

U	120.000
U Prime	136.000
Z-Value	-.302
P-Value	.7630
Tied Z-Value	-.597
Tied P-Value	.5506
# Ties	2

The final objective measure considered was that of secondary task errors. Only those subjects exposed to the secondary task could be measured on this attribute (n=16) and only one independent variable could be considered (System Response Time). Table 13 provides the descriptive statistics for the effect of system response time on secondary task error rate. Error rate was calculated by dividing the number of errors by the number of screens exposed to the subject. In the case of the 2 second condition, this averaged 38 screens, while the 8 second condition averaged 47 screens. Similar to the dialing error rate measure, the distribution for the secondary task error rate failed to conform to the normal distribution as determined by a Kolmogorov-Smirnov One-Sample Test. As such, the non-parametric test, Mann-Whitney U was performed, yielding the results displayed in Table 14.

TABLE 13

System Response Time Effect on Secondary Task Error Rate Means, Experiment 1

	Count	Mean	Std. Dev.	Std. Error
Two Seconds	8	.089	.088	.031
Eight Seconds	8	.036	.056	.020

TABLE 14

Mann-Whitney U Test for Effect of System Response Time on Secondary Task Error Rate, Exp. 1

U	14.500
U Prime	49.500
Z-Value	-1.838
P-Value	.0661
Tied Z-Value	-1.903
Tied P-Value	.0570
# Ties	3

An analysis of the mean response times for each trial reflect a tendency for practice effects as evidenced by the means table in Table 15, analysis of variance values in Table 16, and the plot in Figure 1. Figure 1 also presents the means with error bars reflecting one standard deviation in either direction of each mean. Trials had a clear significant impact upon response time. The longer drop off for the final trial is explained by subjects' pressing the keys prior to the completion of the voice prompt, resulting in a measured response time of 0 seconds. The practice effects analyses were calculated with n=30 due to the loss of two data points during collection. This loss did not effect an other analyses as mean response times were calculated without the missing trial data point. However, when examining practice effects over trials, this absence could not be ignored.

TABLE 15

Practice Effects Means, Experiment 1

	Count	Mean	Std. Dev.	Std. Error
Trial 1	30	4.215	2.603	.475
Trial 2	30	3.583	2.143	.391
Trial 3	30	3.043	.986	.180
Trial 4	30	2.663	1.102	.201
Trial 5	30	2.409	.904	.165
Trial 6	30	1.100	1.344	.245

TABLE 16

ANOVA for Practice Effects, Experiment 1

Source	df	Sum of Squares	Mean Square	F-Value	P-Value
Subject	29	218.939	7.550		
Trials	5	171.784	34.357	19.913	.0001
Trials * Subject	145	250.172	1.725		

Dependent: Trials

TABLE 17

Subjective Scale Items: Descriptive Statistics

Scale	Mean	Std Dev.
Overall reactions to the system		
1. Not Satisfying...Satisfying	6.53	.842
2. Difficult...Easy	6.56	1.10
3. Useless...Useful	6.62	.707
4. Not Interested...Interested	6.38	1.16
5. Slow...Fast	6.16	1.19
Overall reactions to the use of voice by the system		
6. Unhelpful...Helpful	6.53	.842
7. Unpleasant...Pleasant	6.44	1.19
Overall reactions to the instructional label on the phone		
8. Unhelpful...Helpful	6.85	.37
9. Unclear...Clear	6.63	1.13
Could find the correct keys (buttons) on the phone		
10. With Difficulty...With Ease	6.90	.39
11. Slowly...Fast	6.47	1.19
Number of keys (buttons) needed to access the system		
12. Unreasonable...Reasonable	6.72	.77
Ease of Access		
13. Inefficient...Efficient	6.81	.47
14. Inconvenient...Convenient	6.66	1.1
Stopping the system after activation (e.g. the final trial)		
15. Difficult...Easy	6.85	.37
16. Inconvenient...Convenient	6.69	1.1
A Personal Identification number security device to start the call		
17. Not Desirable...Desirable	4.75	2.27
18. Useless...Useful	5.22	1.88
Would you be willing to pay $5 per month for this service?	Yes=24 No=8	

Table 18 provides information for an Item Analysis of the questionnaire scale items using a Chi-Square Analysis. The analysis treats the data as a multinomial experiment, as described by Mendenhall and Beaver (1991). Each scale, consisting of seven possible responses (1-7) theoretically could result in an equal distribution of responses across the seven items. The "expected" values (32 subjects / 7 possible responses = 5.47) were compared to the "observed" values in a Chi Square Analysis, and displayed in Table 18. The p-values indicate that scales 1 through 16 are significant at least at the .005 confidence level and scale 18 is significant at the .01 level. Appendix L provides bar charts for each of the scales depicting the distribution of responses.

Examining the tables in Appendix L allows the reader to view the positively skewed shape of the distributions, violating the normal distribution assumption necessary for parametric analyses. For this reason, the non-parametric test, Mann-Whitney-U was used to probe for significant differences in subjective attitudes between the two system response time conditions. No analyses resulted in a significant difference at the .05 level.

TABLE 18

Item Analysis of Questionnaire Data

	Item							Chi-Square Value	P-Value
	1	2	3	4	5	6	7		
Scale									
1	0	0	0	2	1	7	22	85.71	0.005
2	1	0	0	0	0	8	23	97.97	0.005
3	0	0	0	0	4	4	24	101.03	0.005
4	0	1	0	2	1	7	21	76.52	0.005
5	0	1	2	0	4	8	17	49.83	0.005
6	0	0	1	0	1	9	21	82.65	0.005
7	1	0	0	1	0	9	21	82.65	0.005
8	0	0	0	0	0	5	27	132.98	0.005
9	1	0	0	0	1	4	26	119.85	0.005
10	0	0	0	0	1	1	30	165.36	0.005
11	1	0	0	1	0	8	22	88.34	0.005
12	0	0	1	0	0	5	26	121.60	0.005
13	0	0	0	0	1	4	27	131.23	0.005
14	1	0	0	0	0	5	26	121.60	0.005
15	0	0	0	0	0	5	27	132.98	0.005
16	1	0	0	0	0	4	27	131.23	0.005
17	6	1	2	3	4	6	10	12.19	0.1
18	3	1	1	4	4	10	9	17.01	0.01
Totals	15	4	7	13	22	109	406		
Averages	0.83	0.22	0.39	0.72	1.22	6.06	22.6		

The first section of the questionnaire asked the subjects for basic background information regarding their age, education, exposure to "high-tech" products, telephone habits, and 911 habits. Table 19 provides a summary of this data subdivided by independent variables (see Appendix F for complete questionnaire items). Education is based on 12 = high school degree, with each additional number representing one more year of college, etc. Product Usage is based on the average number of "high-tech" products the respondent is familiar with or owns (maximum = 14). The "Have 911?" item provides the number of respondents whose home is covered by 911 service. Phone

Usage is based on the average number of times the respondent uses the phone per day. The "Phone Services" item is based on the average number of phone services the subjects use (e.g. call waiting, call trace, repeat dialing, etc.) with a maximum of 8. The "911 Usage" item reflects the average number of times the subjects have called 911 or the local emergency number in their lifetime.

TABLE 19

Summary of Questionnaire Background Information

	Cognitive loading		No Cognitive loading		Average
	2 Sec. SRT	8 Sec. SRT	2 Sec. SRT	8 Sec. SRT	
Age	65.50	69.75	74.00	71.00	70.06
Education	15.38	14.38	16.25	16.25	15.56
Product Usage	5.75	5.38	4.50	4.38	5.00
Have 911?*	8	8	8	8	8
Phone Usage	7.13	4.69	5.13	4.88	5.45
Phone Services	.75	.63	.38	1.00	.69
911 Usage?	2.62	.75	.63	.25	1.06

* All values given are means except "Have 911" displays the number of respondents who indicated that they have 911 service in their area. SRT = System Response Time.

An examination of the distribution curves for many of the subjective and objective measures indicates the normal distribution assumption required by parametric tests is not present, necessitating a non-parametric test. The Spearman Rank-Order Correlation Coefficient test is one of the most popular tests for measuring the association between two variables and has a power-efficiency of 91% of its powerful parametric counterpart (Siegel and Castellan, 1988). Tables 20, 21, and 22 display the non-parametric association analyses which yielded significant correlations at the .05 level.

TABLE 20

Spearman Rank-Order Correlation Coefficient for Initial Response Time, Average Response Time

Sum of Squared Differences	1005.500
Rho	.816
Z-Value	4.542
P-Value	<.0001
Rho corrected for ties	.816
Tied Z-Value	4.541
Tied P-Value	<.0001
# Ties, Initial Time	1
# Ties, Average Time	2

TABLE 21

Spearman Rank-Order Correlation Coefficient for Initial Response Time, Subscribed Phone Services

Sum of Squared Differences	7465.500
Rho	-.368
Z-Value	-2.051
P-Value	.0403
Rho corrected for ties	-.498
Tied Z-Value	-2.772
Tied P-Value	.0056
# Ties, Services	3
# Ties, Initial Time	1

TABLE 22

Spearman Rank-Order Correlation Coefficient for Calling Errors, Subscribed Phone Services

Sum of Squared Differences	2176.000
Rho	.601
Z-Value	3.347
P-Value	.0008
Rho corrected for ties	.312
Tied Z-Value	1.739
Tied P-Value	.0821
# Ties, Services	3
# Ties, Calling Errors	2

The results of the brief memory test following the other items on the questionnaire were highly positive as nearly all subjects perfectly recalled which keys were necessary to press in order to activate the specific emergency service branches.

Experiment 2

Objective Measures. Three primary objective measures were used to determine the success of system usage by children. The first two, as in Experiment 1 were the initial and average response times. The third was the number of dialing errors from the total possible number of dialing errors.

Table 23 provides the descriptive statistics for the effect of system response time on average user response time. Table 24 provides the analysis of variance summary table for system response time (with the same two levels as in Experiment 1) and the dependent measure of average response time. As the p-values indicate, system response time resulted in no significant effect on this dependent measure. However, as will be discussed later, there was a tendency for the longer system response time interval resulting in longer response times on average.

TABLE 23

Effect of System Response Time on Average Response Time Means, Experiment 2

	Count	Mean	Std. Dev.	Std. Error
Two Seconds	8	3.385	1.299	.459
Eight Seconds	8	4.286	1.375	.486

TABLE 24

ANOVA for Average Response Time, Experiment 2

Source	df	Sum of Squares	Mean Square	F-Value	P-Value
SRT	1	3.249	3.249	1.816	.1992
Residual	14	25.047	1.789		

Dependent: Average Time
SRT = System Response Time

A post hoc power analysis for the effect of system response time on average response time revels that 77 subjects would have had to been tested to result in statistically significant differences with an alpha level of .05 and a power level of .80.

The third main objective dependent measure was that of dialing errors. Dialing error rate is calculated the same as in Experiment 1. The descriptive statistics for which are provided in Table 25. Informal examination of the distribution as well as the results of a Kolmogorov-Smirnov One-Sample Test reflect a clear violation of the normal distribution assumption, justifying the use of a non-parametric test. Table 26 displays the results of a Mann-Whitney U Test, indicating no significant difference between system response time conditions on dialing error rate.

TABLE 25

Effect of System Response Time on Dialing Error Rate Means, Experiment 2

	Count	Mean	Std. Dev.	Std. Error
Two Seconds	8	.004	.011	.004
Eight Seconds	8	.005	.009	.003

TABLE 26

Mann-Whitney U Test for Effect of System Response Time on Dialing Errors, Experiment 2

U	28.000
U Prime	36.000
Z-Value	-.420
P-Value	.6744
Tied Z-Value	-.620
Tied P-Value	.5351
# Ties	2

Table 27 provides the descriptive statistics for the trial effect on response time. Table 28 presents analysis of variance values indicating response times significantly decreased over the 6 trials, similar to those of the elderly. Figure 2 presents a plot of this data.

TABLE 27

Practice Effects Means, Experiment 2

	Count	Mean	Std. Dev.	Std. Error
Trial 1	16	4.954	2.293	.573
Trial 2	16	4.321	2.154	.539
Trial 3	16	3.663	1.811	.453
Trial 4	16	3.184	1.435	.359
Trial 5	16	3.748	1.590	.398
Trial 6	16	3.293	2.860	.715

TABLE 28

ANOVA for Practice Effects, Experiment 2

Source	df	Sum of Squares	Mean Square	F-Value	P-Value
Subject	15	167.698	11.180		
Trial	5	35.801	7.160	2.424	.0431
Trial * Subject	75	221.579	2.954		

Dependent: Trial

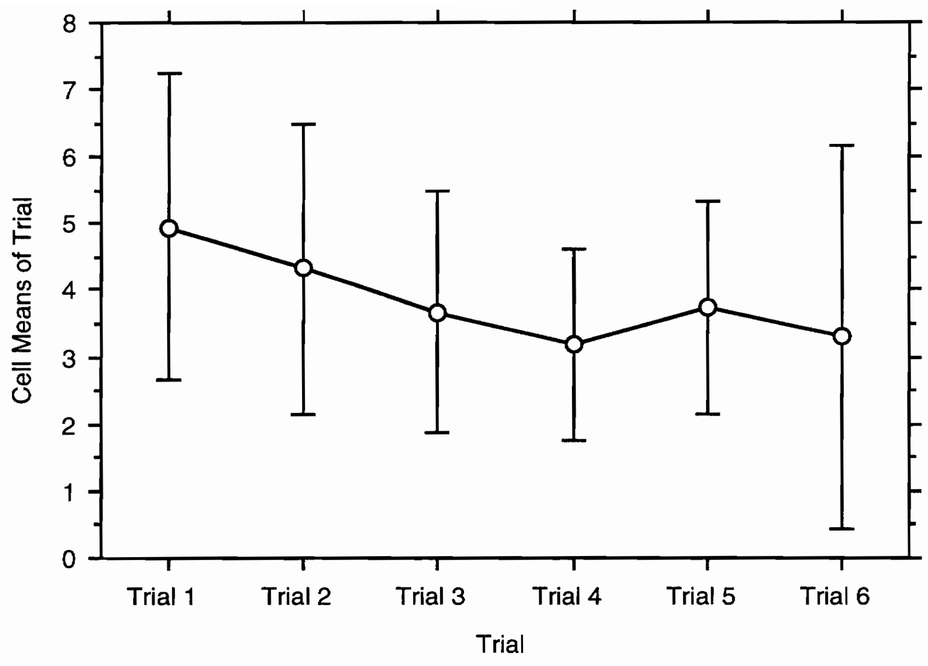


FIGURE 2

Plot of Practice Effects, Experiment 2

<u>Subjective Measures.</u> The main subjective measures for Experiment 2 were taken from the scale items in the questionnaire (see Appendix G). Each scale item consisted of a statement followed by a scale similar to that of the questionnaire for the elderly, with the exception that only three points delineated the scale, as opposed to seven for the elderly questionnaire The right column of Table 29 provides the item statement and the three "anchor words". A value of 1 was always assigned to the leftmost word, representing the negative pole of the dimension, while the right end, assigned a value of 3, represented the positive pole of the dimension.

TABLE 29

Subjective Scale Items: Descriptive Statistics, Experiment 2

Scale	Mean	Std Dev.
How simple was the system to use?		
Hard, In Between, Easy	2.63	.50
How was the time dialing and the voice message?		
Slow, In Between, Fast	1.81	.75
Could you read the big label? (pointed to first word)		
No, A little, Yes	1.18	.54
Could you understand the voice on the telephone?		
No, A Little, Yes	2.94	.25
Did the telephone work fast or slow?		
Slow, In Between, Fast	2.13	.81
Stopping the phone in the middle (e.g. the last time) was		
Hard, In Between, Easy	2.88	.34

To examine any differences between system response time conditions for the subjective scale items, a Mann-Whitney U non-parametric test was employed, which according to Siegel and Castellan (1988), is used for independent samples of ordinal data. None of the six scale items produced a significant difference with an alpha level of .05.

Combined Data

Because the two experiments were conducted in a highly similar manner, it may be of interest to make comparisons between the two data sets where appropriate. Table 30 provides the descriptive statistics for the age effect on initial response time. Table 31 provides analysis of variance information for a between-subjects comparison of the effects of age and system response time on initial response time. Figure 3 similarly shows the plot for the effect of age on initial response time. Age had a mildly significant effect, causing longer initial response times for the children.

A post hoc Power Analysis for the effect of age on initial response time indicates that nearly 68 subjects would be required to obtain a significant difference, given the obtained difference magnitude, an alpha level set at .05 and a power level of .80.

TABLE 30

Effect of Age on Initial Response Time Means, Combined Data

	Count	Mean	Std. Dev.	Std. Error
Old	16	3.682	1.843	.461
Young	16	4.954	2.293	.573

TABLE 31

ANOVA for Initial Response Time, Combined Data

Source	df	Sum of Squares	Mean Square	F-Value	P-Value
Age	1	12.941	12.941	2.990	.0941
Residual	30	129.842	4.328		

Dependent: Initial Time

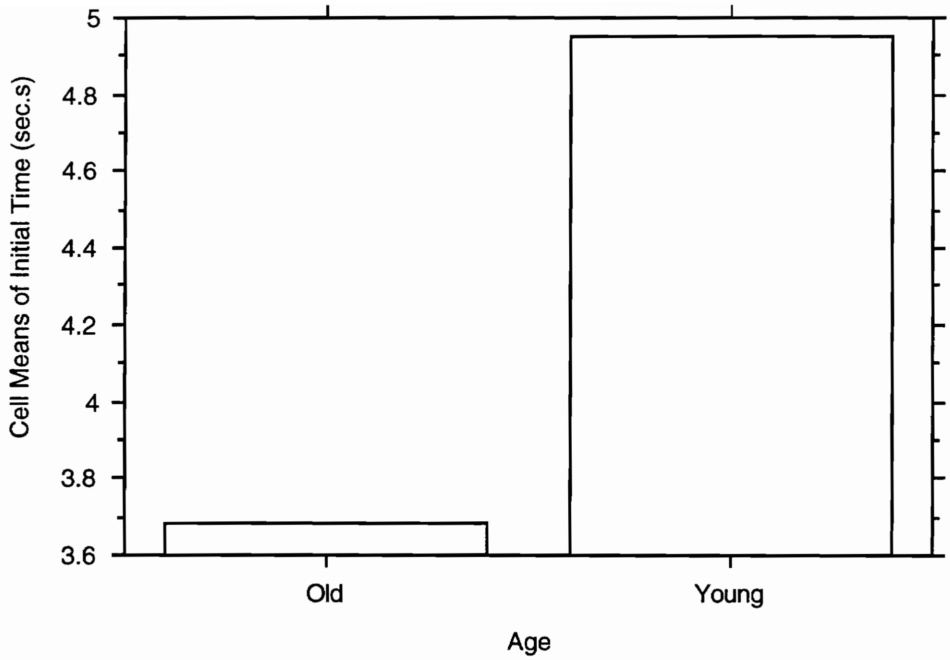


FIGURE 3

Plot of Age Effect on Initial Response Time, Combined Data

Table 32 displays the descriptive statistics for the age effect on average response time. Table 33 similarly displays the descriptive statistics for the system response time effect for the combined age groups on average response time. The effects of age and system response time on average response time were analyzed with a 2x2 between-subjects analysis of variance in Table 34. Age had a highly significant effect, while system response time produced an insignificant effect. Figure 4 displays the mean average response times for the effect of age.

TABLE 32

Effect of Age on Average Response Time Means, Combined Data

	Count	Mean	Std. Dev.	Std. Error
Old	16	2.668	.825	.206
Young	16	3.836	1.373	.343

TABLE 33

Effect of System Response Time on Average Response Time, Combined Data

	Count	Mean	Std. Dev.	Std. Error
Two Seconds	16	2.979	1.165	.291
Eight Seconds	16	3.524	1.334	.333

TABLE 34

ANOVA for Average Response Time, Combined Data

Source	df	Sum of Squares	Mean Square	F-Value	P-Value
Age	1	10.916	10.916	8.706	.0063
SRT	1	2.382	2.382	1.899	.1791
SRT * Age	1	1.012	1.012	.807	.3767
Residual	28	35.109	1.254		

Dependent: Average Time

SRT = System Response Time

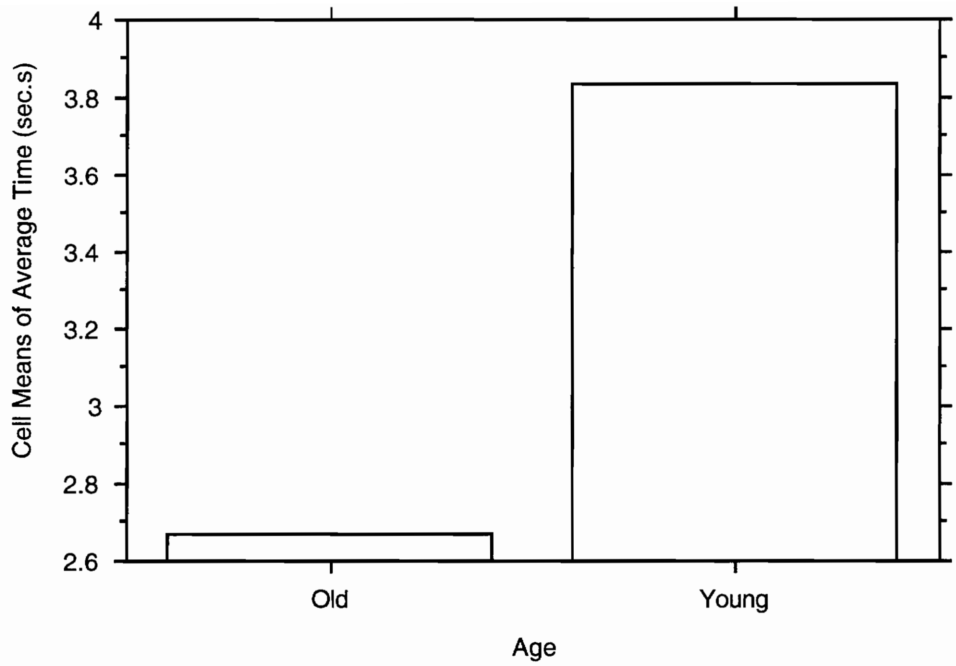


FIGURE 4

Plot of Age Effect on Average Response Time, Combined Data

A post hoc Power Analysis for the effect of system response time on average response time indicates that 75 subjects would had to have been tested to result in statistically significant differences with an alpha level of .05 and a power level of .80.

Tables 35 and 36 display the descriptive statistics for the effects of age and system response time on the final objective measure, dialing error rate. Examination of the dialing error rate distribution by way of the Kolmogorov-Smirnov One-Sample Test suggests considerable departure from the normal distribution. For this reason, the non-parametric test, Mann-Whitney U was used to assess the effects of System Response Time and Age on Dialing Error Rate. The results of this analysis are displayed in Tables 37 and 38, and indicate that Dialing Error Rate was not significantly affected by either age or system response time at the .05 level.

TABLE 35

Effect of Age on Dialing Error Rate Means, Combined Data

	Count	Mean	Std. Dev.	Std. Err.
Old	16	0.000	0.000	0.000
Young	16	.016	.033	.008

TABLE 36

Effect of System Response Time on Dialing Error Rate Means, Combined Data

	Count	Mean	Std. Dev.	Std. Err.
Two Seconds	16	.005	.021	.005
Eight Seconds	16	.010	.028	.007

TABLE 37

Mann-Whitney U Test for Effects of System Response Time on Dialing Error Rate, Combined Data

U	120.000
U Prime	136.000
Z-Value	-.302
P-Value	.7630
Tied Z-Value	-.597
Tied P-Value	.5506
# Ties	2

TABLE 38

Mann-Whitney U Test for Effects of Age on Dialing Error Rate, Combined Data

U	104.000
U Prime	152.000
Z-Value	-.905
P-Value	.3657
Tied Z-Value	-1.789
Tied P-Value	.0736
# Ties	2

DISCUSSION

Consistent with the previous section, the discussion of results will focus on each of the two experiments in terms of both objective and subjective data, and then the comparisons that can be drawn between the two studies. In all cases in which statistics resulted in a failure to reject the null hypothesis, the language will be couched in terms of "no significant differences" for purposes of clarity and consistency.

Experiment 1

Objective Measures. The primary objective measures included the initial and average response times, dialing error rate, and secondary task error rate. The two independent variables, system response time and cognitive loading had no significant effect on either initial or average response time. While not statistically significant, these results do provide interesting information in regard to the ease with which the system was used.

A post hoc Power Analysis revealed that increasing the sample size to a reasonable level would not have resulted in significant differences for the effect of cognitive loading on initial response time, despite the low p-value associated with this difference. Approximately 155 subjects would have been necessary to achieve significant differences considering the size of the effect, a reasonable power level (.80) and an alpha level of .05. Therefore, because the cognitive loading condition did not significantly slow the use of the system, the conclusion can be drawn that the system was designed in a simple manner, facilitating speedy access.

Alternatively, such a result opens several new questions, precluding such a positive conclusion. One hypothesis is that the secondary task was sufficiently simple and easy, avoiding the detrimental effects of a true diversion. Because the subjects were given the opportunity to practice the secondary task for ten cycles prior to interacting with the primary telephone task, they were able to achieve a level of proficiency approaching that of an automatic routine, which has been shown to produce smaller interference effects than tasks requiring controlled processing as discussed by Korteling (1994) and Schneider and Fisk (1982).

The goal of the secondary task was to gain an understanding of whether or not the system could successfully be operated in an attention-demanding environment, such as would result in the event of a true emergency situation. While the creation of such an environment is unethical and measuring its impacts practically impossible, the secondary task proved to be a best guess, showing that the system could be successfully operated in that specific situation, a positive result, despite doubts of its strength.

As mentioned earlier, system response time also had no significant impact on average user response time. Butler (1983) showed that user response time increased somewhat with longer system response times. The differences between the two system response time conditions were not statistically significant, disallowing support for Butler's finding. Another explanation for the weak differences incurred by the secondary task lies in the nature of studying the behavior of elderly adults. Due to the numerous physical and psychological degenerations occurring at drastically different ages, motor and intellectual functioning in older adults fluctuates considerably. However, post hoc Power Analyses revealed that even a much larger sample size would not dampen individual differences and allow the experimental variables to cause more noticeable effects. Other more appropriate measures of the effect of system response time are discussed in the section on subjective measures.

The third objective measure, dialing error rate, was not normally distributed according to a Kolmogorov-Smirnoff One-Sample Test, and therefore was analyzed with the non-parametric Mann-Whitney U test. The results of which revealed no significant effect of cognitive loading on dialing error rate. Three subjects in the cognitive loading condition committed one dialing error each, while no dialing errors were committed in the non-cognitive loading condition. Overall few errors were committed, due in part to familiarity with the telephone key-pad and the instructions to prioritize the dialing over the secondary task performance. In the case of the three subjects who committed the errors, one simply forgot to press the pound key on the last trial, until reminded. The other two subjects pressed incorrect service keys (e.g. pressed "1" rather than "2") and did not realize the error until the voice confirmation message did not conform to their expectation. These two errors were left uncorrected at the suggestion of the experimenter, as the prototype was not prepared at that point to recover from an error with the "*" key as designed.

The effects of system response time on dialing error rate were insignificant (one error was made in the two second condition and one was made in the eight second condition). Such a result does not conform to the findings of Weiss, et al. (1982) who found that for a process control task, longer system response times tended to decrease the number of errors committed. Perhaps, given a larger number of trials, and therefore a greater opportunity to commit errors, the result would be similar to Weiss, et al. Another indication of the effects of system response time on performance lies in the secondary task error rate.

The number of errors committed on the secondary task in the cognitive loading condition was recorded to determine if different dual-task strategies were employed by the subjects and if the independent variable of system response time had any noticeable effects. Again, the data was not normally distributed, and therefore was analyzed with the Mann-Whitney U test which indicated that the different system response time conditions did not result in different error rates. This conforms to Butler's (1983) finding of no task performance decrements due to longer system response times.

Similarly, nearly all errors were kept to a minimum, and failed to indicate different patterns of performing the two simultaneous tasks. An interesting phenomenon, however, was the observation that nearly every subject who committed at least one error, made an error at the onset of the primary telephone task, at the end of the ten practice cycles for the secondary task. Such an occurrence can be explained by the abrupt change in task requirements at the onset of the primary task for which a new motor and cognitive activity must be initiated following an auditory prompt. After the initial, unaccustomed modification, the subjects usually eased into the rhythm of the requirements of the dual-task situation.

The final observation from the objective data was a clear propensity for shorter response times over the six trials. A within-subjects ANOVA of the six trials indicates pronounced practice effects, $\underline{F}(29, 5) = 19.91$, $\underline{p} = .0001$. Subjects were able to learn which key on the key-pad corresponded to the appropriate service over time, and probably grew comfortable with the task itself. Removing the effects of practice may have been accomplished through repeated practice until an asymptotic level had been reached, but this would have precluded the determination of first-time access-speed, an important measure for an infrequently-used system of such importance. Incidentally, initial response time was significantly correlated with average response time as evidenced by the Spearman Rank-Order Correlation Coefficient. Such a finding can be interpreted as a general tendency for the practice effect to be somewhat constant in magnitude across subjects. In other words, the measure for the average response time decreased by approximately the same amount from the initial response time for all subjects.

The practice effects finding may have implications for actual use of the system in terms of maintaining a fresh familiarity with the system. It can be suggested that users periodically activate the system to maintain knowledge and comfort with the system but this would doubtfully effect performance in the rare event of an emergency situation. It was mentioned in the Results section that some subjects were able to press the keys on the keypad prior to completion of the voice prompt, causing the software to measure a response time of 0 seconds, which contributed to the low mean response time of the final trial.

Despite the strong evidence indicating statistically significant practice effects, the finding must be considered in light of practical significance. It is clear that in nearly every access opportunity in a real situation, a user would not be provided with the luxury of practicing the key presses, nulling the usefulness of such an effect. Yet the initial access time should be reviewed to determine if it was satisfactory. The average initial response time (4.21 seconds) was approximately 3 seconds less than the sixth and final response time (1.1 seconds). Accessing the service in less than 5 seconds appears to be rather fast, but should be compared against the time typically taken to access the emergency responders through current means (usually either a 3 or 7 digit number). Although emergency responders should be contacted as quickly as possible, it is highly unlikely that a 3 second savings would have any impact on an emergency situation, especially in light of the many other factors that influence the speed with which emergency responders arrive at a scene.

Non-Parametric tests of association were conducted on many of the questionnaire items and the objective measures. Only two significant corrleations on the Spearman Rank-Order Correlation Coefficient resulted: Between the initial response time and the number of subscribed phone services and between dialing error rate and the number of subscribed phone services. Those subjects who subscribed to more phone services tended to have lower initial response time and dialing error rates. One plausible explanation for this result may be the familiarity or degree of comfort those subjects had in dialing unique telephone key-sequences and lengths (two-digit sequences as opposed to standard seven- and ten-digit sequences). This added experience allowed these subjects to dial faster with fewer errors. Vigorous conclusions ought not be made from these results as the number of calling errors committed was small and the range of subscribed services was narrow (0-3).

Subjective Measures. The chief means for collecting subjective attitudes and data for the emergency telephone system were through a questionnaire, comprised of two main sections. The first section asked for basic background information which was used to characterize the subject pool in terms of demographics and technology familiarity, and the scale section, used to collect subjects' opinions and attitudes about various characteristics of the system. In general, the subject pool was highly educated (mean = 3.5 years of college, with a range of eighth grade to Ph.D.), elderly (mean=70 years old, with a range of 60 to 81), and represented a mix of sexes (17 females, 15 males). All subjects lived in an area covered by 911 service. The sample had little to moderate exposure to "high-tech" home consumer products (mean = 5 out of 14 possible items), leading to the conclusion that this segment of the population is not the most technically advanced nor in a market class which favors purchasing the latest electronic products. In regard to phone services offered by the local telephone company (e.g. call forwarding, call return, etc.), the sample had little experience as the average number of subscribed services was .69 out of 8 possible. Such a result suggests that this group of people are reluctant to pay ancillary surcharges for telephone services.

The most important segment of the questionnaire contained 18 scale items as described in earlier sections (see Appendix F). The first 16 scales all produced mean responses between 6 and 7 on a 7-point scale, demonstrating either highly favorable opinions and attitudes toward the system or a faulty questionnaire. Making the assumption that the questionnaire could detect true differences in attitudes toward the system, the results will be discussed in this light. The overall reactions to the system were favorable, across all four conditions. The system response time and cognitive loading did not bias subjects' attitudes as the non-parametric Mann-Whitney U test indicated no such difference.

The scale item concerning system speed yielded no differences between the two system response time conditions, suggesting a suitable tolerance for the longer system response time, consistent with the findings of Murray and Abrahamson (1983), but orthogonal to the findings of Bergman, et al. (1981). Shneiderman (1987) offers a reasonable explanation for the discrepancy between these findings. Shneiderman claims that previous experience plays a large role in users' tolerance for system delays. Since none of the subjects had experience with the specific task domain, they had no expectations on this dimension. Additionally, typical phone calls often have considerable lags prior to connection, which may have influenced subjects' tolerance for the task at hand.

The use of voice was favored in the system as a confirmation message and the instructional labeling scheme was considered helpful. The keys on the key-pad were considered easy to locate and an appropriate number to dial for system access, which was thought to be convenient and efficient. Finally, stopping the system in the event of an accidental actuation was considered to be an easy task.

The final two scale items concerned the use of a personal identification number (PIN) for system access, and yielded relatively negative responses. This result is not surprising, especially when the PIN feature was described in more detail, as the subjects recognized its drawbacks, confirming the system design decision to avoid the entry of an additional number prior to system access.

An item analysis was performed on the questionnaire data, the results of which are depicted in Table 17 and Appendix L. The Item Analysis involved using a Chi-Square test statistic to compare a uniform distribution ("expected values") to observed values. Scales 1-16 were highly significant, with p-values less than .005. Scales 17 and 18 were significant at the .1 and .01 level, respectively. These results suggest that the scales were able to detect clear attitudes of the subjects, with the exception of Scale 17. Insignificant differences, conversely, indicate that the scales were not successful in measuring attitudes. The Chi-Square provides no evidence that certain item values were significantly more prevalent than others (e.g. 7 chosen more often than 6 on the scale), however, glancing through the plots in Appendix L reveals that a value of 7 was indeed chosen much more than the other values. Scales 17 and 18, with lower p-values, most likely display a more uniform distribution of responses due to the nature of their questions (See discussion on PIN Number usage above).

At the end of the questionnaire, subjects were asked whether they would be willing to pay $5 per month for such a service. The fact that 75% indicated that they would pay, is an indication of the value the subjects placed on the service. Of those who said they would not be willing to pay, several suggested that given a more serious medical condition in the household, the service would have considerably more appeal.

In summary, the subjective measures demonstrated that the phone was received positively by the subjects. The interface was considered adequate, the method of access and instructional scheme appropriate, and the longer system response time condition produced no impact on user satisfaction.

Experiment 2

Objective Measures. Literature and discussions with parents verify that children are first taught what actions to take in the event of an emergency (e.g. what telephone number to call) at approximately 5 years old, prior to kindergarten. Many of these children learn this important information in a pre-school or day-care program. The purpose of testing children this age was to substantiate whether or not they could indeed make the call, much the same as a 911 call, and understand what the call means. Additionally, the experiment attempted to provide data to judge the effectiveness of the labeling scheme, and evaluate the effect of system response time. The same objective measures as in Experiment 1 were used to gauge the influence of the independent variable, system response time.

System response time had no effect on initial response time, average response time, or dialing errors, although there was a tendency toward longer average response times for the 8 second system response time condition. One predicted result was that the longer system response time condition would cause confusion about whether or not the call actually went through, resulting in increased dialing errors. This was not the case, however, suggesting that the 8 second system response time is a suitable system characteristic.

A distinct practice effect was observed over the six trials, $F(15, 5) = 11.18$, $p=.043$, as was the case for the elderly. Again the practical significance of this finding should be examined. The average initial response time was 4.95 seconds, roughly 1.5 seconds slower than the sixth trial. Such a difference is not likely to be an important factor in the speedy arrival of the emergency responders. Additionally, an access time of less than 5 seconds is fairly fast, but should be compared against the access times for current emergency telephone numbers prior to making vigorous conclusions.

Subjective Measures. As in Experiment 1, the primary means of obtaining subjective attitudes and opinions from subjects was by way of a questionnaire. During pilot testing, it was learned that children of pre-school age have difficulty discriminating more than three levels of a particular construct. For this reason, the questionnaire contained only three levels from which to make a decision regarding the given statement. While administering the questionnaire, the experimenter felt that some of the questions or the concept of a scale were not sufficiently well-understood. Children also may have been subject to a recency or primacy effect. Such an effect causes the responder to choose either the first or last option presented. Based on this feeling, the scale item data will not be emphasized in forming the basis for conclusions about the system.

Despite such skepticism, one item did tend to reflect differences between the two system response time conditions. Item two, regarding the length of time between dialing and the confirmation message tended to have more negative responses from those subjects who experienced the 8 second system response time. Such a result suggests that children recognize a lengthy time lag, and may have negative feelings for such a system characteristic. Because the system response time did not inhibit performance, the conclusion can be made that a longer delay, though disliked, does not impact the actions critical to successful system operation. The other five scale items did not reflect consistent differences between the two system response time conditions.

Combined Data

Subjective measures cannot be compared between the two age groups due to drastically different questionnaire item formats and delivery methods. The objective measures of response time and error rate can, however, be compared, given a notable difference in the experimental protocol. As indicated in Appendix H, I and J, the instructions for the two different age groups differ according to their cognitive abilities. This precaution was obviously necessary in order to convey the instructions effectively to each subject, and doubtfully imposed a profound influence on user response time and error rate.

The results of a 2x2 between-subjects ANOVA suggest that age had a statistically significant effect on average response time. Age differences were not, however, evident on the initial response time data. A post hoc Power Analysis for the effect of age on initial response time showed that approximately 68 subjects would be necessary for a statistically significant difference. In effect, the difference between age groups for average response time was greater than the difference for initial response time. One explanation for this finding may lie in the ability of the adults to master the task and feel more comfortable with the testing situation more so than the children. Similarly, the adults may have been as disoriented as the children at the start of the study. The data from the practice effects analysis show that the effect of practice was greater for the adults than the children, though both differences were statistically significant.

In general, however, older subjects tended to respond faster than their pre-school counterparts. This finding may not seem consistent with the literature on the tendency for behavioral slowing in the elderly (Salthouse, 1982). However, a middle age group would be required prior to making such a conclusion. Additionally, as Cerella (1985) argues, the slowing of sensory-motor processes is less severe than the slowing of higher order processes, requiring considerably more complex thought routines. The task of calling the emergency service was a simple routine for elderly adults, and did not require a great deal of thought, especially given the brief instructions prior to activation, their long experience with telephones, and the salient dual label format.

Children, conversely, have limited experience with telephones, cannot read the labeling scheme, and in some cases, were distracted during the instruction briefing. On average the young group used the phone approximately .8 times per week, while the elderly used the phone over 30 times per week, and likely have done so for many years. In terms of reading ability, only one young subject could read the large label on the phone, while all the elderly subjects were literate. As mentioned earlier, the elderly read the instructions silently while the experimenter read aloud. Children, simply listened to the experimenter, with the parent interjecting if the child appeared confused. Finally, the elderly subjects appeared to pay attention adequately during the instructions, while some of the children exhibited the tendency to become distracted during this phase. All these factors may have contributed to a hesitation or reluctance to activate the service. One young subject became frightened just at the point when he was supposed to activate the service, presumably for fear of not knowing what would result. The elderly subjects all understood that the experiment was a ruse and would not actually contact a real emergency responder.

The fact that there was a larger difference between the two age groups for the average response time than the initial response time may reflect the tendency for the elderly subjects to become more comfortable with the system quicker than the children. The children may also have been distracted during the experimental session, resulting in delayed responses across all trials.

System response time effects for the combined data did not result in significant differences for average user response times, F (1, 28) = 1.899, p = .179. Such a low p-value suggests the sample size was not large enough to measure differences. A post hoc Power Analysis, however, indicated that nearly 75 subjects would have been required to elicit statistically significant differences with a power level of .80 and alpha level of .05. This projected sample size is probably too large for a study of this size. Future experiments, however, may be able to resolve this issue more fully.

The final measure, dialing error rate did not reflect a significant difference between the two age groups. Similarly, system response time had no effect on dialing error rate for the combined age group data.

CONCLUSIONS

Several major findings have resulted from the data collected for the first experiment. Given the interface employed, system response time and cognitive loading had no significant effects on initial user response time, average user response time, or dialing error rate. Furthermore, secondary task error rate was not significantly affected by the system response time. In terms of qualitative measures, the scale items suggested that the subjects held favorable opinions for the system, had a high tolerance for system response time, and were willing to pay for the service.

These findings support several claims. First, the user interface, designed with considerable regard to previous relevant literature, was sufficient for older, first-time users. Secondly, the system was considered easy to use, even under conditions of increased cognitive loading. The instructional format was helpful and adequate in making the users comfortable and capable of first-time access. Finally, longer system response times were not considered detrimental by the subjects.

These findings contribute to the literature in a number of ways. The findings relating to system response time tolerance fill a narrow, but valuable gap in understanding human behavior in telephone-based applications. As the current study examined only two points on the temporal continuum for a single application, more research ought to investigate this issue, expanding and updating Miller's (1968) review. Because user response time was not affected by increased cognitive loading, such telephone-based applications making use of touch-tone entry can be considered an adequate means for data entry in similar dual-task environments. More research should, however, investigate various cognitive loading levels, perhaps delving into the workload literature, engaging more realistic secondary tasks. Much could be learned from examining what levels of cognitive loading effect the operation of telephone-based data entry. Finally, the use of elderly participants on a new telephone service has not been reported in the literature, making this study a unique profile of elderly behavior.

The second experiment afforded a number of interesting conclusions as well. As in the first experiment, initial and average response times as well as dialing error rate were not affected by the system response time. The questionnaire scales were shown to present dubious representations of the attitudes of young children.

These findings support several claims. First, it can be concluded that because the dialing error rate was low, children were able to access the emergency services adequately. The instructional interface (including the iconic mapping of emergency service to telephone keypad), therefore, was sufficient for a population of users unable to read. Secondly, the children were able to access the system without increased errors for the longer system response time condition. Although this may have little impact on performance in the real world, the fact that the subjects did not become confused and concerned during the longer system response time interval, indicates that longer system response times are adequate for children of this age.

As in the first experiment, the data from the children compliment the research on telephone usage and applications by providing a glimpse as to how this age group operates telephones for novel applications. Further research could examine children's long-term memory (rehearsed or unrehearsed) for accessing emergency services. The results clearly indicate slower response times for the children than the elderly. To further investigate this relationship, other age groups could be tested, increasing our understanding of behavior slowing with respect to age.

Schumacher (1992) points out that despite the rapid changes in the telecommunications arena, standard touch-tone telephone interfaces will continue to be a ubiquitous, speedy, and easy tool for personal communications and that published work in this area is limited. This research has demonstrated that an alternative emergency service can be designed, enabling fast, error-free access to automated messages through the telephone key-pad. The effects of system response time and increased cognitive-loading on user performance and acceptability were determined. Finally, the usage capabilities and subjective opinions of elderly adults and young children were examined.

REFERENCES

Abacus Concepts (1992). *StatView Manual.* Berkeley, CA: Abacus Concepts.

Aucella, A. F. and Ehrlich, S. F. (1986). Voice messaging: enhancing the user interface based on field performance. In *Proceedings of CHI '86 Human Factors in Computing Systems.* (pp. 156-161). New York, NY: Association of Computing Machinery.

Backman, L. and Molander, B. (1986a). Adult age differences in the ability to cope with situations of high arousal in a precision sport. *Psychology and Aging*, 1(2), 133-139.

Backman, L. and Molander, B. (1986b). Effects of adult age and level of skill on the ability to cope with high-stress conditions in a precision sport. *Psychology and Aging,* 1(4), 334-336.

Bailey, J. E. and Pearson, S. W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management Science*, 29(5), 530-545.

Bailey, R. W. (1989). *Human Performance Engineering.* Englewood Cliffs, NJ: Prentice Hall.

Bergman, H., Brinkman, A., and Koelega, H. S. (1982). System response time and problem solving behavior. In *Proceedings of the Human Factors Society 25th Annual Meeting.* (pp. 749-753). Santa Monica, CA: Human Factors Society.

Birren, J. E. and Cunningham, W. (1985). Research on the psychology of aging: principles, concepts, and theory. In J. E. Birren and K. W. Schaie (Ed.s) *Handbook of the Psychology of Aging, Second Edition.* (pp. 3-34) New York, NY: Van Noorstrand Reinhold Company.

Blanchard, H. E., Lewis, S. H., Ross, D., and Cataldo, G. (1993). User performance and preference for alphabetic entry from 10-key pads: Where to put the Q and Z? In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 225-229). Santa Monica, CA: Human Factors Society.

Blumberg, H. H., De Soto, C. B., and Kuethe, J. L. (1966). Evaluation of rating scale formats. *Personnel Psychology,* 19, 243-259.

Bowe, F. G. (1991). National survey on telephone services and products: The views of deaf and hard of hearing people. *American Annals of the Deaf, 136* (3), 278-283.

Booher, H. D. (1975). Relative comprehensibility of pictorial information and printed words in proceduralized instructions. *Human Factors*, *17* (3), 266-277.

Botwinick, J. and Kornetsky, C. (1960). Age differences in the acquisition and extinction of the gsr. *Journal of Gerontology*, 15, 83-84.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32, 513-531.

Butler, T. W. (1983). Computer response time and user performance. In *Proceedings of CHI '83 Human Factors in Computing Systems* (pp. 58-62).New York, NY: Association for Computing Machinery.

Byblow, D. (1990). Effects of redundancy in the comparison of speech and pictorial displays in the cockpit environment. *Applied Ergonomics, 22* (1), 121-128.

Card, S. K., Moran, T. P., and Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin*, 98(1), 67-83.

Chapanis, A. (1973). The communication of factual information through various channels. *Information Storage and Retrieval,* 9, 215-226.

Chin, J. P., Diehl, V. A., and Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the CHI '88 Conference on Human Factors in Computing Systems* (pp. 213-218). New York, NY: Association for Computing Machinery.

Cochran, W. G. (1971) Some consequences when the assumptions for the analysis of variance are not satisfied. In Steger, J. A. (Ed.) *Readings in Statistics for the Behavioral Scientist.* (pp. 330-346). New York, NY: Holt, Rinehart, and Winston, Inc.

Coleman, W. D. and Williges, R. C. (1985). Collecting detailed user evaluations of software interfaces. In *Proceedings of the Human Factors and Ergonomics Society 29th Annual Meeting* (pp. 240-244). Santa Monica, CA: Human Factors Society.

Cushman, W. H. and Rosenberg, D. J. (1991). *Human Factors in Product Design.* Amsterdam: Elsevier Science Publishers B. V.

DeVellis, R. F. (1991). *Scale Development: Theory and Applications.* Newbury Park, CA: Sage Publications.

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.

Durch, J. S. and Lohr, K. N. (Ed.s). (1993). *Emergency Medical Services for Children.* Washington, D.C.: National Academy Press.

Eisdorfer, C. (1967). New dimensions and a tentative theory. *The Gerontologist*, 7, 14-18.

Eysenck, M. W. (1976). Arousal, learning, and memory. *Psychological Bulletin, 83* (3), 359-404.

Finn, R. H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 32, 255-265.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling amplitude of movement. *Journal of Experimental Psychology*, 67, 103-112.

Gould, J. D.. and Boies, S. J. (1983). Human factors challenges in creating a principle support office system: the speech filing system approach, *ACM Transactions on Office Information Systems, 1* (4), 273-298.

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

Halstead-Nussloch, R. (1989). The design of phone-based interfaces for consumers. In *Proceedings of the CHI '89 Conference on Human Factors in Computing Systems* (pp. 347-352) New York, NY: Association for Computing Machinery.

Harper, B. D. and Norman, K. L. (1993). Improving user satisfaction: the questionnaire for user interaction satisfaction version 5.5. In *Proceedings of the First Mid-Atlantic Human Factors Conference* (pp. 224-228) Old Dominion University: Human Factors Society, Tidewater Chapter.

Hawkins, J. S., Reising, J. M., Lizza, G. D., and Beachy, K. A. (1983). Is a picture worth 1000 words-- Written or spoken?, In *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 970-972). Santa Monica, CA: Human Factors Society.

Henneman, R. H. (1952). Vision and audition as sensory channels for communication. *The Quarterly Journal of Speech, 38*, 161-166.

Hix, D. and Hartson, H. R. (1993). *Developing User Interfaces: Ensuring Usability Through Product and Process*. New York, NY: John Wiley and Sons, Inc.

Israelski, E. W. (1988). An experimental comparison of user performance with alternative access codes for PBX features. In *Proceedings of the 12th International Symposium on Human Factors in Telecommunications* (pp. 1-7). the Hague, the Netherlands.

Jenkins, G. D. and Tabor, T. D. (1977). A monte carlo study of factors affecting three indices of composite scale reliability. *Journal of Applied Psychology*, 62(4), 392-398.

Kausler, D. H. (1991). *Experimental Psychology, Cognition, and Human Aging, Second Edition.* New York, NY: Springer-Verlag.

Keppel, G. (1991). *Design and Analysis: A Researcher's Handbook.* Englewood Cliffs, NJ: Prentice Hall.

Kohfeld, D. L. (1971). Simple reaction time as a function of stimulus intensity in decibels of light and sound. *Journal of Experimental Psychology, 88* (2), 251-257.

Korteling, J. E. (1994). Effects of aging, skill modification, and demand alternation on multiple-task performance. *Human Factors*, 36(1), 27-43.

Kramer, J.J. (1968). Human factors problems in the use of push-button telephones for data entry. In *The Fourth International Symposium on Human Factors in Telephony* (pp. 241-258).

Madden, J. M. and Bourdon, R. D. (1964). Effects of variations in rating scale format on judgment. *Journal of Applied Psychology,* 48(3), 147-151.

Mahood, G. K. (1971). Human factors in touch-tone data systems. *Bell Laboratories Record, 12* , 345-348.

Meister, D. (1985). *Behavioral Analysis and Measurement Methods.* New York, NY: John Wiley and Sons, Inc.

Mendenhall, W. and Beaver, R. J. (1991). *Introduction to Probability and Statistics.* Boston, MA: PWS-Kent Publishing Co.

Miller, M. A. and Elias, J. W. (1991). Using menus to access computers via phone-based-interfaces. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 235-239). Santa Monica, CA: Human Factors Society.

Miller, P. H., Haynes, V. F., DeMarie-Dreblow, D., and Woody-Ransey, J. (1986). Children's strategies for gathering information in three tasks. *Child Development*, 57, 1429-1439.

Miller, P. H. and Weiss, M. G. (1981). Children's attention allocation, understanding of attention, and performance on the incidental learning task. *Child Development,* 52, 1183-1190.

Miller, R. B. (1968). Response time in man-computer conversational transactions. In *Proceedings of the National Computer Conference* (pp. 268-277). Montvale, NJ: AFIPS Press.

Mitchell, P. P., and Todd, M. (1985). Human interface for system to create lists of phone numbers. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 95-99). Santa Monica, CA: Human Factors Society.

Murray, R. P. and Abrahamson, D. S. (1983). The effect of system delay and delay variability on inexperienced videotex users. *Behaviour and Information Technology*, 2(3), 237-251.

Nugent, W. A. (1987). A comparative assessment of computer-based media for presenting job task instructions. In *Proceedings of the Human Factors Society 31st Annual Meeting* (pp. 696-700). Santa Monica, CA: Human Factors Society.

Nunnally, J. C. (1967). *Psychometric Theory.* New York, NY: McGraw-Hill Book Company.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning.* Urbana, IL: University of Illinois Press.

Osgood, C. E. (1962). Studies on the generality of affective meaning systems. *American Psychologist*, 17(1), 10-28.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach.* New York, NY: Oxford University Press.

Riley, C. A. (1987). Experiences with an integrated voice and text message service. In *Proceedings of the Human Factors Society 31st Annual Meeting,* (pp. 270-274). Santa Monica, CA: Human Factors Society.

Robinson, C. P. and Eberts, R. E. (1985). There's a voice in the sky; but can a picture tell you more? In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 61-65). Santa Monica, CA: Human Factors Society.

Robinson, C. P. and Eberts, R. E. (1987). Comparison of speech and pictorial displays in a cockpit environment. *Human Factors, 29* (1) 31-44.

Roemer, J. M., Pendley, W. L., Stempski, M. O., and Borgstrom, M. C. (1986). Field study of a voice mail system: design and design-process implications. *SIGCHI Bulletin* (18)2 60-61.

Root, R. W and Koster, C. (1986). Experimental evaluation of a mnemonic command syntax for controlling advanced telecommunications services. In *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 809-813). Santa Monica, CA: Human Factors Society.

Salthouse, T. A. (1982). *Adult Cognition: An Experimental Psychology of Human Aging.* New York, NY: Springer-Verlag.

Savage, P. A., Kemp, D. G., and Payne, D. E. (1991). Auditory versus visual presentation of help messages. In *Proceedings of the Human Factors Society 35th Annual Meeting*, (pp. 244-248). Santa Monica, CA: Human Factors Society.

Schneider, W. and Fisk, A. D. (1982). Concurrent automatic and controlled visual search: can processing occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 8(4), 261-278.

Schumacher, R. M. (1992) Phone-based interfaces: research and guidelines. In *Proceedings of the Human Factors Society 36th Annual Meeting,* (pp. 1051-1055). Santa Monica, CA: Human Factors Society.

Serafin, C., Wen, C., Paelke, G., and Green, P. (1993). Car phone usability: A human factors laboratory test. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 220-224). Santa Monica, CA: Human Factors and Ergonomics Society.

Shneiderman, B. (1987). *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Reading, MA: Addison-Wesley Publishing Company.

Siegel, S. and Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences* . New York, NY: McGraw-Hill, Inc.

Simonov, P.V., Frolov, M.V., Evtushenko, V.F., and Sviridov, E.P. (1977). Effect of emotional stress on recognition of visual patterns. *Aviation, Space, and Environmental Medicine, (8),* 856-858.

Shaffer, D. R. (1989). *Developmental Psychology: Childhood and Adolescence, Second Edition.* Pacific Grove, CA: Brooks/Cole Publishing Company.

Spieth, W. (1965). Slowness of task performance and cardiovascular diseases. In A. T. Welford and J. E. Birren (Ed.s), *Behavior, Aging, and the Nervous System* (pp. 366-400). Springfield, Ill: Charles Thomas.

Stern, K. R. (1984). An evaluation of written, graphics, and voice messages in proceduralized instructions. In *Proceedings of the Human Factors Society 28th Annual Meeting* (pp. 314-318). Santa Monica, CA: Human Factors Society.

Strayer, D. L., Wickens, C. D., and Braune, R. (1987). Adult age differences in the speed and capacity of information processing: 2. An electrophysiological approach. *Psychology and Aging, 2* (2) 99-110.

Stuart, R., Desurvire, H., and Dews, S. (1991). The truncation of prompts in phone based interfaces: using tott in evaluations. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 230-234). Santa Monica, CA: Human Factors Society.

The Taylor Group, Inc. (1992). *Market opportunity for the smart stick concept.* Prepared for NYNEX Science & Technology. Greenwich, CT: The Taylor Group, Inc.

Texas Advisory Commission on State Emergency Communications. (1994). *Statewide market analysis 9-1-1 awareness and usage study: children and 9-1-1.* Texas Department of Health.

Thompson, L. W. and Nowlin, J. B. (1973). Relation of increased attention to central and autonomic nervous system states. In L. F. Jarvik, C. Eisdorfer, and J. E. Blum (Ed.s) *Intellectual Functioning in Adults: Psychological and Biological Influences* (pp. 107-124). New York, NY: Springer Publishing Company, Inc.

Virzi, R. A. (1991). A preference evaluation of three dialing plans for a residential, phone-based, information service. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 240-243). Santa Monica, CA: Human Factors Society.

Weiss, S. M., Boggs, G., Lehto, M., Shodja, S., and Martin, D. J. (1982). Computer system response time and psychophysiological stress II. In *Proceedings of the Human Factors Society 26th Annual Meeting* (pp. 698-702). Santa Monica, CA: Human Factors Society.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance*. Glenview, IL: Scott, Foresman and Company.

Williges, R. C. (1983). Applying the human information processing approach to human / computer interactions. In W. C. Howell and E. A. Fleishman (Ed.s), *Human Performance and Productivity, Volume 2: Information Processing and Decision Making* (pp. 83-119). Hillsdale, NJ: Lawrence Erlbaum Associates.

Williges, R. C. and Williges, B. H. (1982). Modeling the human operator in computer-based data entry. *Human Factors*, 24(3) 285-299.

Woodruff, D. S. (1985). Arousal, sleep, and aging. In J. E. Birren and K. W. Schaie (Ed.s), *Handbook of the Psychology of Aging, Second Edition.* (pp. 261-295). New York, NY: Van Nostrand Reinhold Company.

Woodson, W. E., Tillman, B., and Tillman, P. (1992). *Human Factors Design Handbook* (2nd Ed.). New York, NY: McGraw-Hill, Inc.

Woodsworth, R. S. and Schlossberg, H. (1954). *Experimental Psychology*, New York, NY: Holt, Reinhart, and Winston, Inc.

Young, S. L. and Wogalter, M. S. (1988). Memory of instruction manual warnings: effects of pictorial icons and conspicuous print. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 905-909). Santa Monica, CA: Human Factors Society.

APPENDIXES

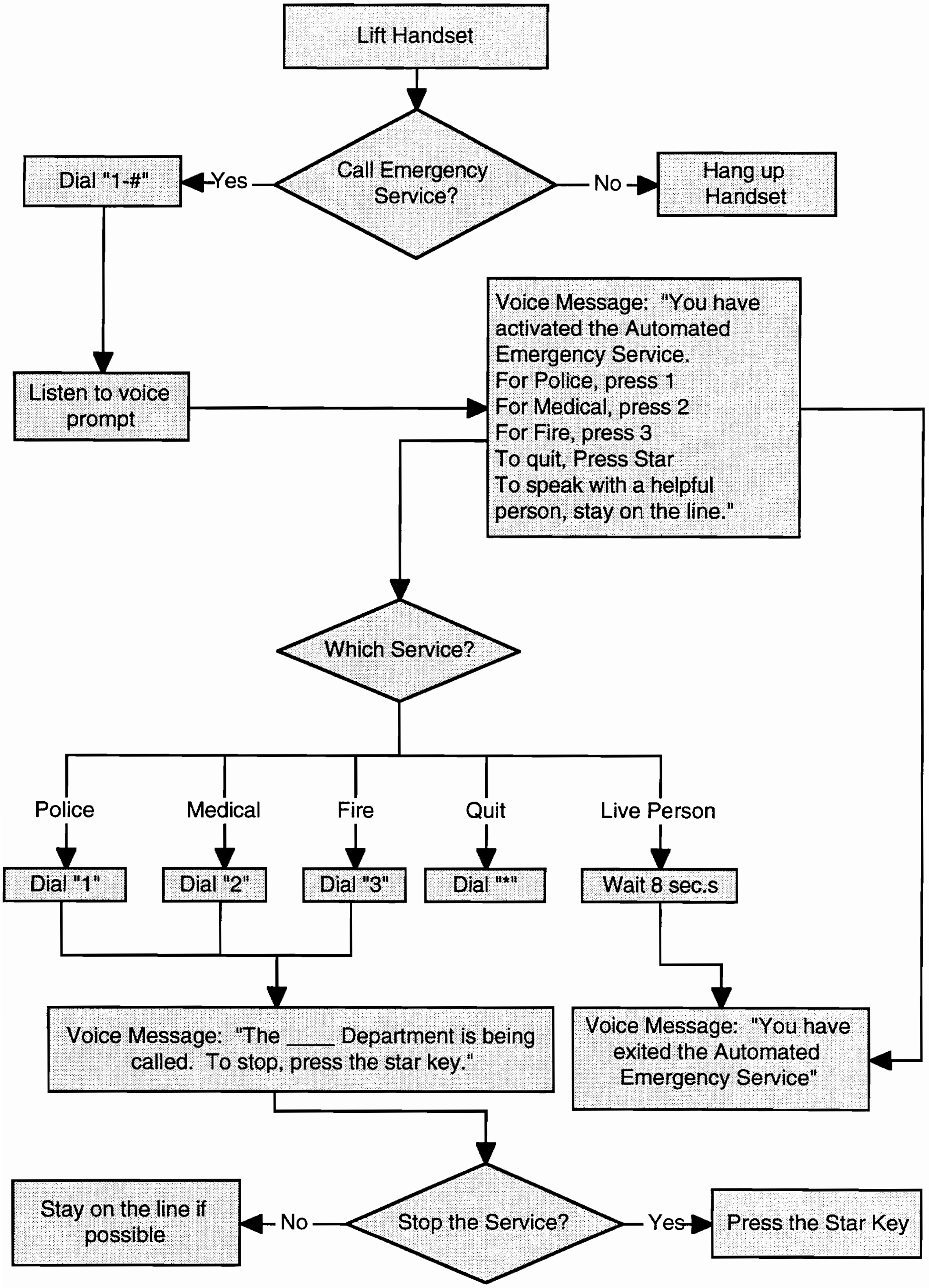
Appendix A - Percentage of Northeast states covered by 9-1-1 Systems

State	Percent
Connecticut	100
Maine	25
Massachusetts	38[1]
New Hampshire	10[1]
New York	80
Rhode Island	100[2]
Vermont	25

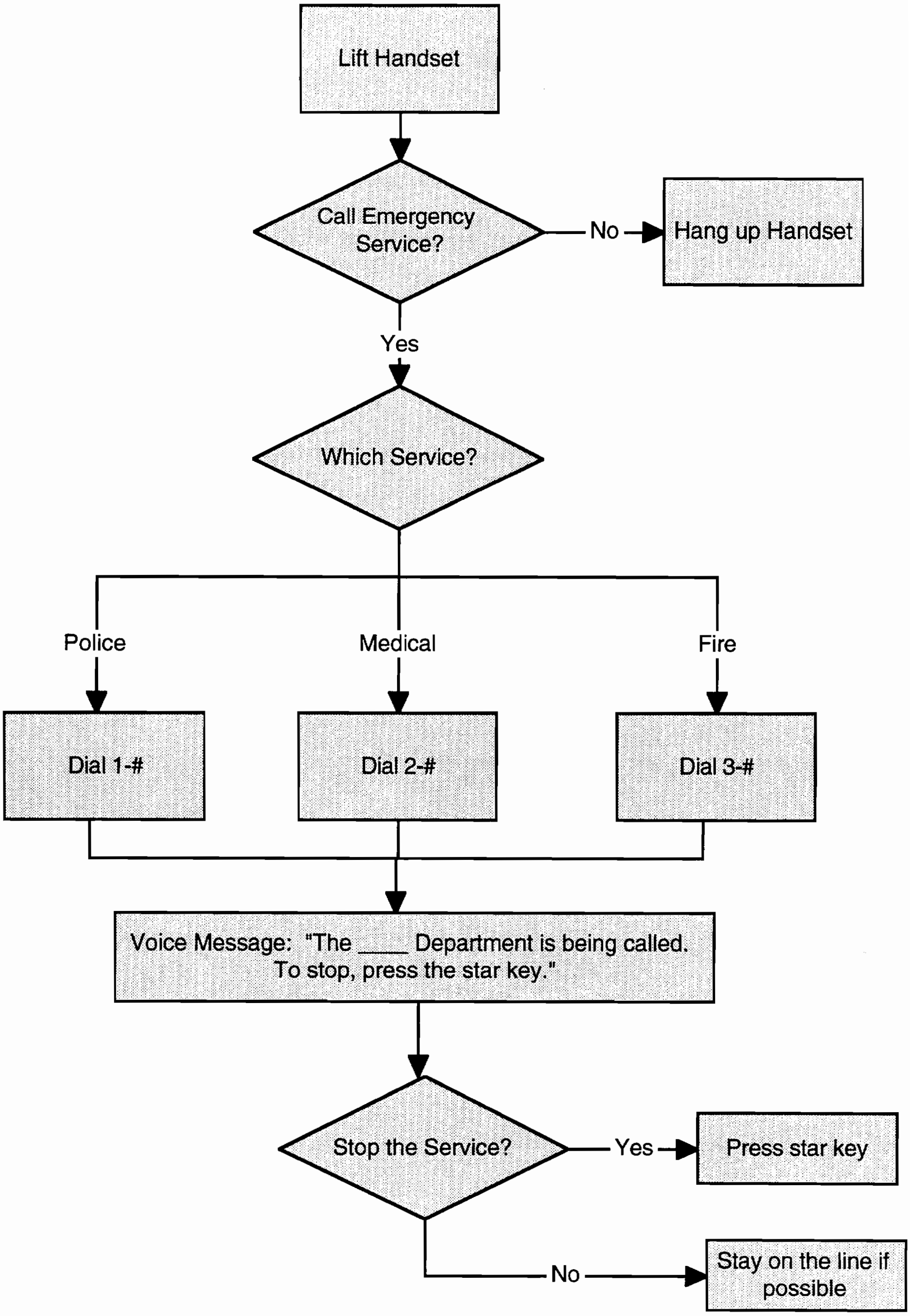
[1] Statewide implementation of enhanced 9-1-1 is underway.
[2] 9-1-1 or similar access number.

Adapted from Emergency Medical Services (1992) as cited in Durch and Lohr (1993)

Appendix B1- Four-Step Version



Appendix B2- Three-Step Version



Appendix C- Fitt's Law Application

Scale Diagram of Key-Pad used in Experiments

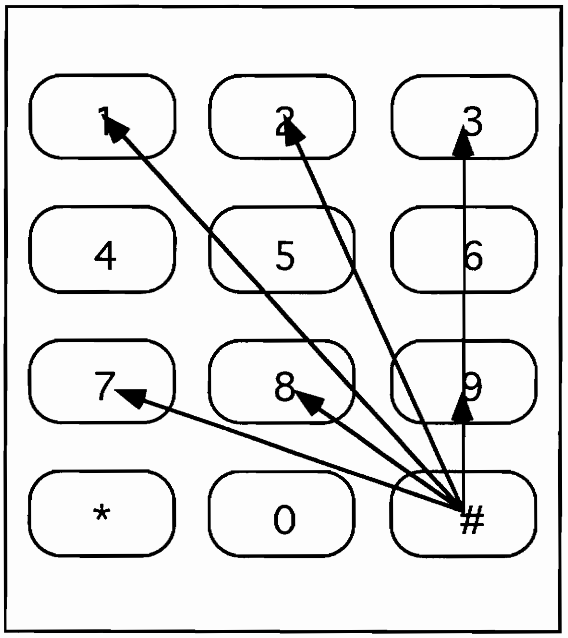


Table 37 depicts a scaled drawing of the actual key-pad used in the experiment. Applying Fitt's Law to this problem will yield differing results for different style key-pads. The above key-pad was used to maintain the experiments' face validity. Each key is drawn to scale (3/4" x 7/16").

Change in Movement Time (MT) $= 100\ [\log_2 (D_1/S_2 + .5) - \log_2 (D_2/S_2 + .5)\ ]$

where I_m = 100 msec.s / bit ; a correction time used in Card, Moran, and Newell (1983), D = the distance between the pound key and the numbered key of interest, and S = the width of the key.

Change in MT between 1 and 7 keys:

$= 100\ [\log_2 (2.875 / 0.750 + 0.500) - \log_2 (2.000 / 0.750 + .5)\ ]$

= 45.25 msec.

Change in MT between 2 and 8 keys:

$= 100\ [\log_2 (2.375 / 0.750 + 0.500) - \log_2 (1.250 / 0.750 + .5)\ ]$

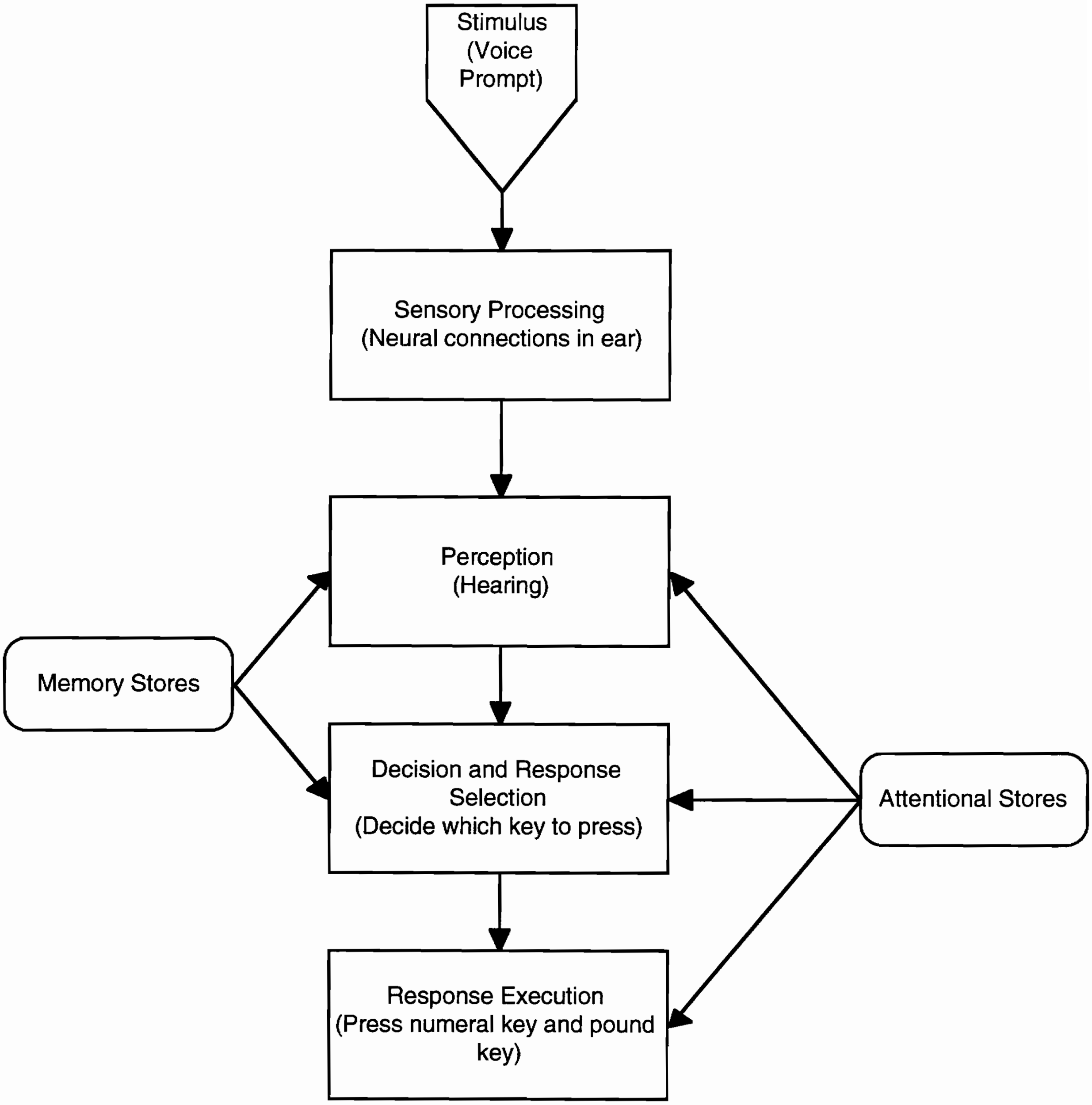
= 75.90 msec.

Change in MT between 3 and 9 keys:

$= 100\ [\log_2 (2.125 / 0.750 + 0.500) - \log_2 (0.750 / 0.750 + .5)\ ]$

= 115.20 msec.

Appendix D- Information Processing Model of Emergency Service Access



Adapted from Wickens (1992)

Appendix E1- Instructional Label for Four-Step Version

In an Emergency:

1. Lift Handset

2. Press 1

3. Press #

4. Listen to the voice prompt
   For **Police**, press **1**
   For **Medical,** press **2**
   For **Fire**, press **3**
   To **Quit**, press *

5. Wait for Authorities

Appendix E2- Instructional Label for Three-Step Version

In an Emergency:

1. Lift Handset

2. For **Police**, press **1**
   For **Ambulance**, press **2**
   For **Fire**, press **3**

3. Press **#**

4. To Quit, press *

5. Wait for Authorities

Additional Template Label:

Police	Medical	Fire
1	2	3
4	5	6
7	8	9
*	0	#
Quit		Enter

Appendix F- Participant Reaction Questionnaire, Experiment 1

Background Information:

Please answer the following questions. This information is requested so we may characterize the people who participate in this evaluation as a group but will NOT be used to characterize or identify specific individuals. Your anonymity will be maintained.

<u>Age :</u>

___ 4-10

___ 11-17

___ 18 - 29

___ 30 - 39

___ 40 - 49

___ 50 - 59

___ 60-69

___ 70-79

___ 80-89

___ 90-

<u>Highest year of education attained:</u> (Circle appropriate number)

9, 10, 11, 12 (High school diploma)

13, 14 (Junior college degree)

15, 16 (College degree)

17, 18 (Masters degree)

19, 20 (Doctorate/MD)

Product Usage/Ownership:

Make a check mark next to those products or services that you own or use.

___ VCR
___ Cellular phone
___ Portable house phone
___ Telephone answering machine
___ Big screen TV (27" or more)
___ Compact disc (CD) player
___ Personal computer
___ Video games (e.g., Atari)

___ Cable TV
___ Premium channels on cable TV
___ Catalog shopping by phone
___ Automated Teller Machines (ATMs)
___ Banking by Phone services
___ Computerized card catalog at library

Phone Services:

Do you reside in an area covered by 911 service?

_____ yes _____ no _____ don't know

On the average, how often do you use the **phone at home**? (Specify number of times for the appropriate category)

_____ Times per day
_____ Times per week

Make a check mark next to those phone services that you use or if you do not know what these services are write "DK" .

___Call waiting
___Call forwarding
___3-way calling
___Call trace
___Call return
___Repeat dialing
___Cancel call waiting
___Caller ID

How many times have you called 911 or the local emergency number?

_____ Times in the past year
_____ Times in your lifetime

We now want you to answer the next series of questions in reference to the emergency service in general. We realize that you have only experienced a brief session with the system but we would still like to obtain your reactions. Circle the number which most closely corresponds to your opinion.

Example:

The coffee at breakfast

distasteful						delicious
1	2	3	4	5	6	7

Overall reactions to the system

not satisfying						satisfying
1	2	3	4	5	6	7

difficult						easy
1	2	3	4	5	6	7

useless						useful
1	2	3	4	5	6	7

not interested						interested
1	2	3	4	5	6	7

slow						fast
1	2	3	4	5	6	7

Overall reactions to the use of voice by the system

unhelpful						helpful
1	2	3	4	5	6	7

unpleasant						pleasant
1	2	3	4	5	6	7

Overall reactions to the instructional label on the phone

unhelpful							helpful
	1	2	3	4	5	6	7

unclear							clear
	1	2	3	4	5	6	7

Could find the correct keys (buttons) on the phone

with difficulty							with ease
	1	2	3	4	5	6	7

slowly							fast
	1	2	3	4	5	6	7

Number of keys (buttons) needed to access the system

unreasonable							reasonable
	1	2	3	4	5	6	7

Ease of access

inefficient							efficient
	1	2	3	4	5	6	7

inconvenient							convenient
	1	2	3	4	5	6	7

Stopping the system after activation (e.g. the final trial)

difficult							easy
	1	2	3	4	5	6	7

inconvenient							convenient
	1	2	3	4	5	6	7

The following feature is currently not part of the system. This purpose of this question is to find out your interest for a security device so that no one else could have access to your personal information.

A personal identification number security device to start the call

not desirable						desirable
1	2	3	4	5	6	7

useless						useful
1	2	3	4	5	6	7

Would you be willing to pay $5 per month for this service?

Yes _______ No _______

Memory Test:

1. What keys must be pressed to access the system?

2. In what order must they be pressed?

Any other comments about the system you wish to share:

Appendix G- Participant Reaction Questionnaire, Experiment 2

Background Information:

Please answer the following questions to help us learn a little about how you liked the telephone system.

Age : ______

Pre-School Level: ______

Phone Services:

Do you reside in an area covered by 911 service?

_____ yes _____ no _____ don't know

How often do you use the phone at home?

_____ Times per day
_____ Times per week

How many times have you called 911 or the local emergency number?

_____ Times in the past year
_____ Times in your lifetime

We now want you to answer the next questions about the emergency service in general. We know that you have only tried it a short time, but we still want your opinion. Circle the number which most closely corresponds to your opinion.

Example:

The juice at breakfast

bad	in between	good
1	2	3

Overall reactions to the telephone calls

hard	in between	easy
1	2	3

slow	in between	fast
1	2	3

Could you read the big label? (point it out)

no	a little	yes
1	2	3

Could you understand the voice on the telephone?

no	a little	yes
1	2	3

Stopping the system in the middle (e.g. the last time) was

hard	in between	easy
1	2	3

Memory Check:

1. What keys must be pressed to access the system?

2. In what order must they be pressed?

Anything else about the system you wish to tell me?

Appendix H- Participant Instructions, Experiment 1 (No Cognitive Loading)

Participant Instructions: Experiment 1a

Welcome!

Please read the following instructions as I read them aloud. If you have any questions, please hold them until the end.

You have been asked to participate in a study testing an alternative system to the familiar 911 emergency service. The tasks required of you are in no way a measure of your abilities or intelligence, but rather, an indication of how usable the system is in general. We value your input and efforts in helping to determine a good design for the system.

The system being tested allows you to activate an automated, pre-recorded message about yourself that would be sent to the appropriate authorities in the event of an emergency. The recording would be personalized to provide information about your household which would enable authorities to respond more efficiently and allow you to attend to the situation yourself without speaking to anyone on the phone.

An example of a personalized message would be the following: "This is an automated call originating from the home of Bill and Sue Mathews at 410 Fuller Avenue in Christiansburg, Virginia. Sue is hearing impaired..." or "This is an automated call originating from the home of the Roger Smith Family at 102 Red Road in Blacksburg, Virginia. The children's rooms are on the third floor. Access to the house can be gained easily through the garage or the side door..."

To activate the system, you need only press a few specified keys on the telephone key-pad. Once the system has been activated, you can leave the room and attend to the emergency, confident that the message is being communicated to the appropriate authorities. In the event that you wish to cancel the message, there is a specific key to stop it.

Every household would have the ability to access three different versions of the service, each corresponding to one of the three primary emergency service branches: Police department, Medical Services, and Fire Department.

In this study you will be asked to respond to seven different emergency situations by calling the appropriate service. A recorded voice will be played over the speaker phone of a standard telephone instructing you to call a specific emergency service. Call as quickly as you can without making a mistake. After a few seconds, a voice will play over the phone telling you that you have reached a specific emergency service. Following the recording, you will make the call, automatically sending off the message. Following the successful completion of a call, you will be instructed to hang the phone up, and another recorded voice will be played for which you are to respond similarly.

After six trials, you will be asked to activate the system one last time. In this trial, you will be asked to stop the system as if you had accidentally called the service.

When finished, you will be given a questionnaire about the service. Your opinions and thoughts about your exposure to the system are very important to us. At this time, any remaining questions you may have about the study or the service will be answered and you will be paid for your time.

Any questions?

Appendix I- Participant Instructions, Experiment 1 (Cognitive Loading)

Participant Instructions: Experiment 1b

Welcome!

Please read the following instructions as I read them aloud. If you have any questions, please hold them until the end.

You have been asked to participate in a study testing an alternative system to the familiar 911 emergency service. The tasks required of you are in no way a measure of your abilities or intelligence, but rather, an indication of how usable the system is in general. We value your input and efforts in helping to determine a good design for the system.

The system being tested allows you to activate an automated, pre-recorded message about yourself that would be sent to the appropriate authorities in the event of an emergency. The recording would be personalized to provide information about your household which would enable authorities to respond more efficiently and allow you to attend to the situation yourself without speaking to anyone on the phone.

An example of a personalized message would be the following: "This is an automated call originating from the home of Bill and Sue Mathews at 410 Fuller Avenue in Christiansburg, Virginia. Sue is hearing impaired..." or "This is an automated call originating from the home of the Roger Smith Family at 102 Red Road in Blacksburg, Virginia. The children's rooms are on the third floor. Access to the house can be gained easily through the garage or the side door..."

To activate the system, you need only press a few specified keys on the telephone key-pad. Once the system has been activated, you can leave the room and attend to the emergency, confident that the message is being communicated to the appropriate authorities. In the event that you wish to cancel the message, there is a specific key to stop it.

Every household would have the ability to access three different versions of the service, each corresponding to one of the three primary emergency service branches: Police department, Medical Services, and Fire Department.

In this study, you will be given the opportunity to call an experimental system. While calling, however, you will be asked to view a computer screen which will display a series of vertical bars. Every few seconds, a new number of bars will be displayed on the screen. You are to simply hold up the number of fingers for each bar that you see. After performing this simple task a number of times, a speaker on the telephone will instruct you to call a specific emergency service.

In this study you will be asked to respond to seven different emergency situations by calling the appropriate service. A recorded voice will be played over the speaker of a standard telephone instructing you to call a specific emergency service. Following the recording, you will make the call, automatically sending off the message. Call as quickly as you can without making a mistake. After a few seconds, a voice will play over the phone telling you that you have reached a specific emergency service. Following the successful completion of a call, you will be instructed to hang the phone up, and another recorded voice will be played for which you are to respond similarly.

Throughout the trials, you will be requested to continue monitoring the computer screen by holding up the number of fingers for each bar displayed. Try to do your best at both tasks, without making mistakes dialing the phone.

After six trials, you will be asked to activate the system one last time. In this trial, you will be asked to stop the system as if you had accidentally called the service.

When finished, you will be given a questionnaire about the service. Your opinions and thoughts about your exposure to the system are very important to us. At this time, any remaining questions you may have about the study or the service will be answered and you will be paid for your time.

Any questions?

Appendix J- Participant Instructions, Experiment 2

Participant Instructions: Experiment 2

Welcome!

My name is Jon, and I am building a new kind of telephone system. The reason I have asked you to come here is to help me see if the new telephone will work properly. If you ever have any questions, please stop me, and I will answer them for you.

Do you know how to call the police or fire department or ambulance if there is an emergency at home? < Check if 911 is known>

The new telephone I am building will be similar to 911. If there is an emergency at home, instead of calling 911, with this phone you would do it a little differently.

Let's pretend that you were home alone and saw smoke coming from the neighbor's house. Normally you would pick up the telephone and dial 911. But, if you had this telephone, you would pick up the phone and dial the number 3 and then this button (show #). If you wanted to talk to the police, you would pick up the phone and dial the number 1 and then this button (show #). If you wanted to call the ambulance, you would dial the number 2 and then this button (show #). Do you have any questions?

Once you have dialed the number you want, just wait and listen. A voice will talk to you telling you that you have reached the emergency service. This is good.

When you are ready to begin, a voice will come from the speaker on the phone and tell you to call either the police, fire or ambulance. You then lift the phone up and press the numbers to call that service as fast as you can without making a mistake. A voice will talk to you telling you that you have reached the emergency service. This is good. Afterwards, another voice will tell you to hang up. Follow the instructions and hang the phone up. A few seconds later, another voice will tell you to call a different service. Just lift up the phone and press the right buttons. This will happen six times. Keep following the instructions.

When you are done, I will ask you to call the police one more time and then stop the call like you made a mistake by pressing the "Quit" button (star). I will explain this more carefully at that time.

When finished, I will ask you a couple questions about how you liked the service.

Do you have any questions?

Appendix K- Post Hoc Power Analysis Calculations

Effect of Cognitive Loading on Initial Response Time, Experiment 1:

$$\hat{\omega}_A^2 = \frac{(a-1)(F-1)}{(a-1)(F-1)+(a)(n)}$$

$$\hat{\omega}_A^2 = \frac{(2-1)(1.436-1)}{(2-1)(1.436-1)+(2)(8)}$$

$$\hat{\omega}_A^2 = .0265$$

where $\hat{\omega}_A^2$ = relative magnitude for treatment effect, A

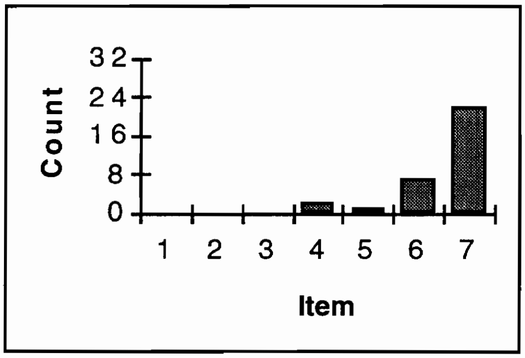
$$\Phi_A^2 = n\frac{\hat{\omega}_A^2}{1-\hat{\omega}_A^2}$$

where Φ_A = Statistic used in computing power in Pearson-Hartley Charts

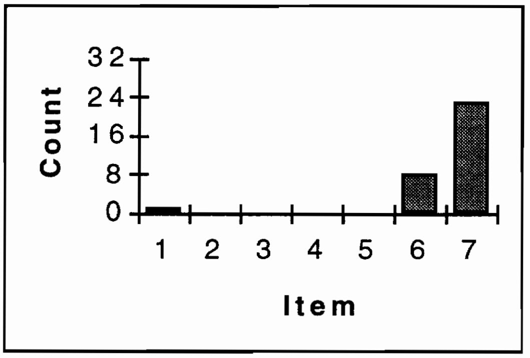
$$\Phi_A = \sqrt{155}\sqrt{\frac{.0265}{1-.0265}}$$

$\Phi_A = 2.05$, using Pearson-Hartley Charts in Keppel, 1991, and maintaining $\alpha = .05$ and $\beta = .80$, 155 subjects needed

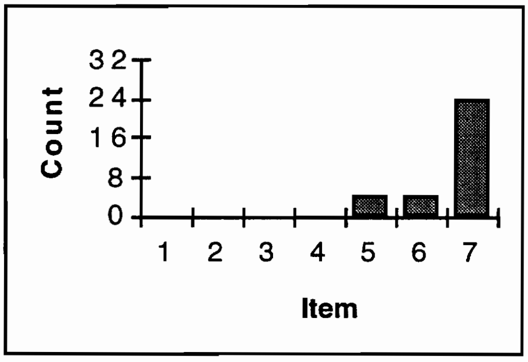
Appendix L- Questionnaire Item Analysis Plots, Experiment 1



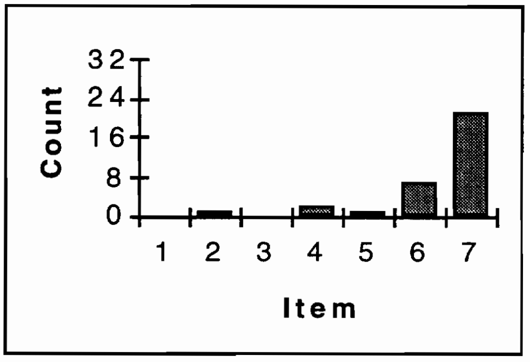
Scale 1



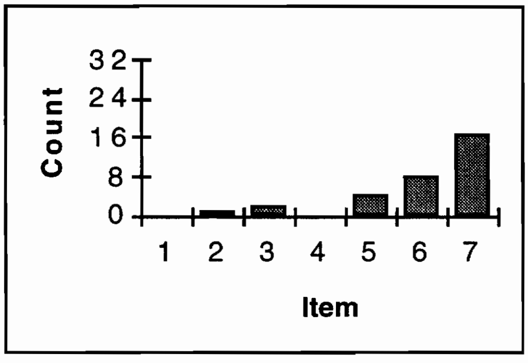
Scale 2



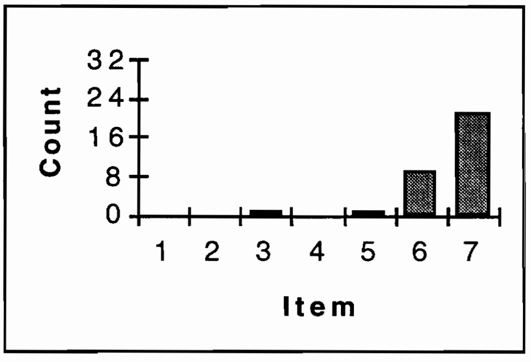
Scale 3



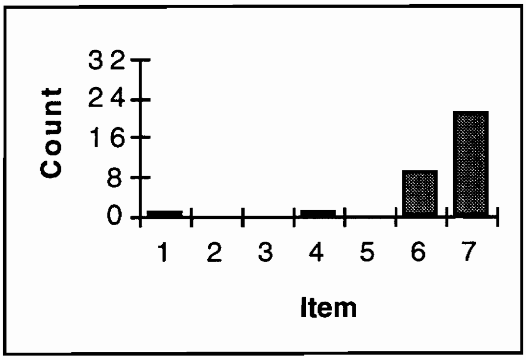
Scale 4



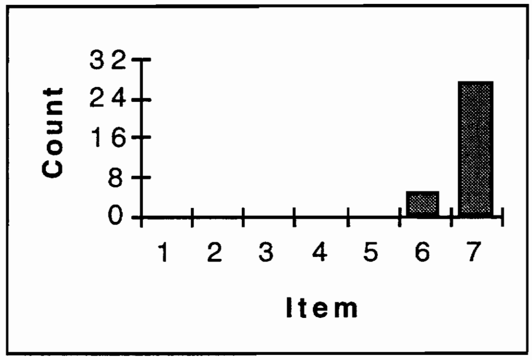
Scale 5



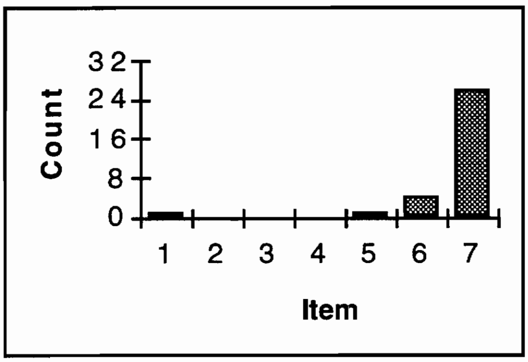
Scale 6



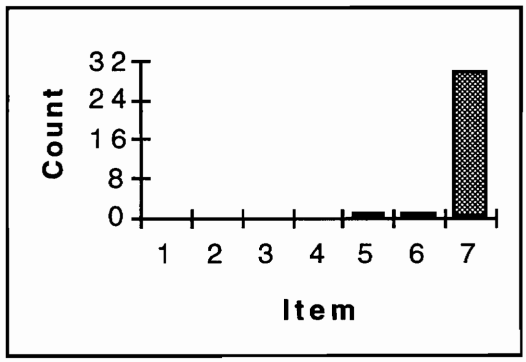
Scale 7



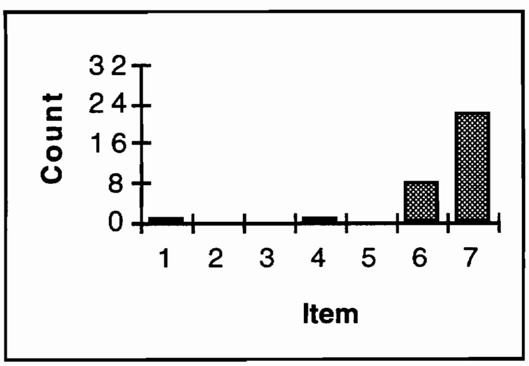
Scale 8



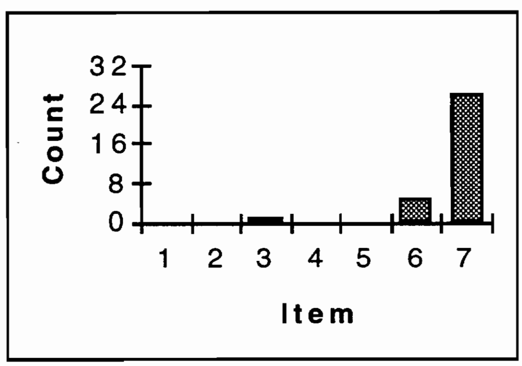
Scale 9



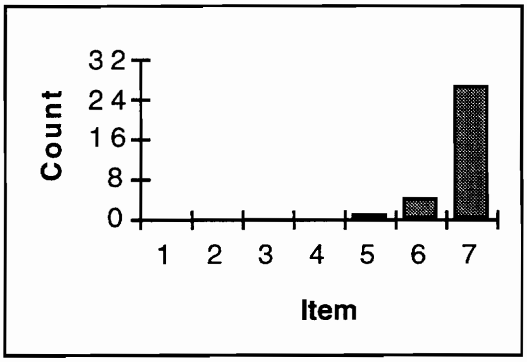
Scale 10



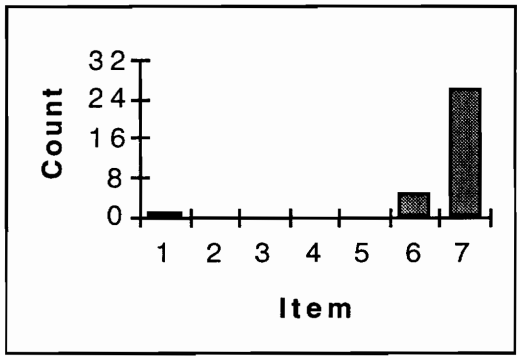
Scale 11



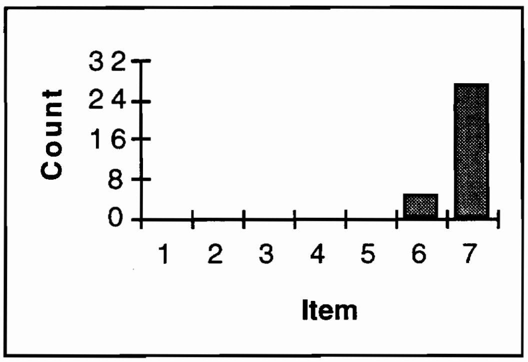
Scale 12



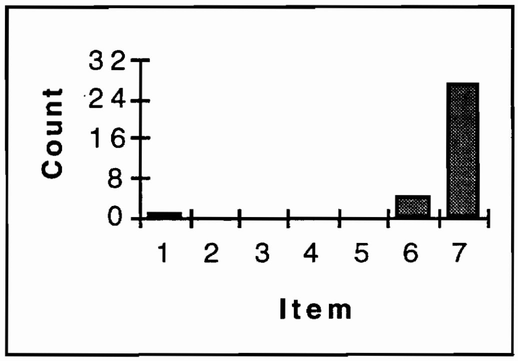
Scale 13



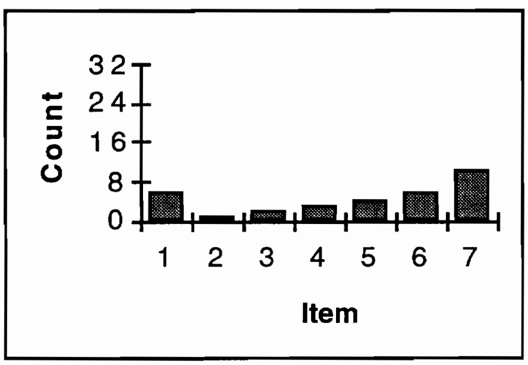
Scale 14



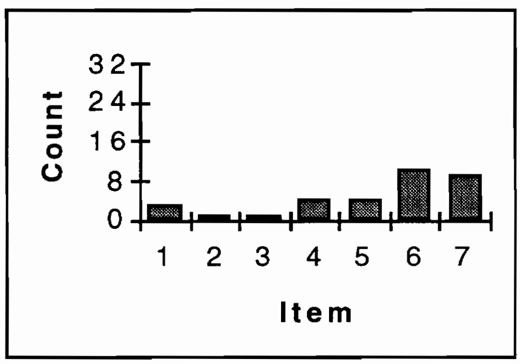
Scale 15



Scale 16



Scale 17



Scale 18

Appendix M- Informed Consent, Experiment 1

Informed Consent Form (Adults)

Emergency Services Experiment, Spring 1994
Human Factors Engineering Center
Environmental and Safety Laboratory

Subject # ______

You have been asked to participate in a study examining your abilities and preferences in regard to a emergency telephone service. The experiment will be conducted in the Environmental and Safety Laboratory of Whittemore Hall at Virginia Tech or at the Blacksburg Baptist Church on 550 North Main Street, Blacksburg. You will be required to participate in one experimental session of approximately one half-hour. You will be paid five dollars for the session.

The results of this study will be strictly confidential. At no time will the data be given to anyone other than the principle investigators without your written consent. The information you provide during the experiment will have your name removed and only a subject number will identify your data. There are no apparent risks to you associated with participation in this study. In the event you wish to terminate your participation for any reason, you will be paid for the time you have provided to the study.

The information from this research may be used for scientific and educational purposes. It may be presented at scientific meetings and/or published and republished in professional journals or books, or used for any other purposes which Virginia Tech and the Industrial and Systems Engineering department considers proper in the interest of education, knowledge, or research. This research has been approved by the Industrial and Systems Engineering department Internal Review Board and by the Institutional Review Board of Virginia Tech.

Please confirm your understanding of the following statements before signing the informed consent form.

- I have read and understand the above description of the study.
- I have had the opportunity to ask questions and have all of them answered.
- I hearby acknowledge the above and give my voluntary consent for participation in this study.
- I understand that if I choose to participate in this study, I may withdraw at any time without penalty.
- I understand that should I have any questions about this research and its conduct, I should contact any of the following:

Graduate Student: Jonathan Kies	231-9087
Primary Researcher: Dr. Dennis Price	231-5635
IRB, ISE: Dr. Robert Beaton	231-5936
Chair, IRB Research Division: Ernest R. Stout	231-9359

Subject's Signature: ________________________________ Date: ___________

Appendix N- Informed Consent, Experiment 1

Informed Consent Form (Children)

Emergency Services Experiment, Spring 1994
Human Factors Engineering Center
Environmental and Safety Laboratory

Your child is being asked to participate in a study examining the "user-friendliness" of an emergency telephone service. The experiment will be conducted at the Environmental and Safety Lab in Whittemore Hall at Virginia Tech. Your child will be asked to participate in one experimental session of approximately twenty minutes. Standard procedure at Virginia Tech is to pay each participant $5 per hour (or fraction thereof), in this case, however, the child will be given a Virginia Tech T-Shirt.

The results of this study will be strictly confidential. At no time will the data be given to anyone other than the principle investigators without your written consent. The information you provide during the experiment will have your child's name removed and only a subject number will identify the data. There are no apparent risks associated with participation in this study. In the event your child feels uncomfortable or wishes to terminate participation for any reason, he/she will be paid for the time provided to the study. Additionally, a parent of other familiar adult may be present at all times during the study.

The information from this research may be used for scientific and educational purposes. It may be presented at scientific meetings and/or published and republished in professional journals or books, or used for any other purposes which Virginia Tech and the Industrial and Systems Engineering department considers proper in the interest of education, knowledge, or research. This research has been approved by the Institutional Review Board of Virginia Tech.

Please confirm your understanding of the following statements before signing the informed consent form.

- I have read and understand the above description of the study.
- I have had the opportunity to ask questions and have all of them answered.
- I hearby acknowledge the above and give my voluntary consent for my child's participation in this study.
- I understand that if I choose for my child to participate in this study, I may withdraw at any time without penalty.
- I understand that should I or my child have any questions about this research and its conduct, I should contact any of the following:

Graduate Student: Jonathan Kies	231-9087
Primary Researcher: Dr. Dennis Price	231-5635
Chair, IRB Research Division: Ernest R. Stout	231-9359

Child's Name: ____________________________

Parent or Guardian's Signature: ____________________________ Date: __________

VITA

Jonathan K. Kies was born March 4, 1970 in Summit, New Jersey. He is currently a Masters student in Industrial and Systems Engineering at Virginia Polytechnic Institute and State University. He received a B. A. in Psychology from Miami University in Oxford, Ohio in 1992. He has worked summers for NYNEX Science & Technology in White Plains, New York. His interests include human factors issues related to telecommucniations, computer networks, and human-computer interface design.

Jonathan K. Kies