ABSTRACT

The Effect of Response Modality on Task Performance When Using an Interactive Voice Response System for Older and Younger Adults

by

Nicole T. Howie

The purpose of this experiment was to gain a greater understanding of how set-level compatibility affects task performance of older adults while using an Interactive Voice Response (IVR) System. Set-level compatibility effects occur when differences in response time are obtained due to characteristics shared between a stimulus and response set. The current study used banking and healthcare IVR tasks to investigate whether manipulating set-level compatibility would differentially affect performance between younger (18 - 39 years old) and older (65 years and older) adults. As expected, subjects performed better in the high set-level compatibility conditions. However, no interaction between set-level compatibility and age was found, which can possibly be attributed to task differences, IVR design and real world context of the task. Overall, the primary contributions of this research are the design guidelines that were created to develop an IVR that suppressed age effects.

Acknowledgments

My utmost gratitude goes to my advisor, Dr. Michael Byrne, for his continued patience, support, and guidance. I would also like to thank my committee members, Dr. Philip Kortum and Dr. Jessica Logan for their advice as well as for their evaluation of this thesis and their time.

My appreciation also extends to my laboratory colleagues. Dr. Kristen K. Greene words cannot express how thankful I am for all of your help. Thanks also to Bryan Campbell for your constant encouragement and stats knowledge.

Thanks to Scott Wilson and Marc Middlekauff of Plum Voice. I would not have been able to complete this research without the use of QuickFuse.

Finally, I would like to acknowledge family, especially my husband, Dee, who has allowed me to realize my dream of going back to school and my son, Titus for surrounding me with love.

TABLE OF CONTENTS

List of Figures

List of Tables

INTRODUCTION

The elderly population is growing rapidly. The U.S. Census Bureau estimates that by the year 2050, there will be 88.5 million people 65 and older in the population (U.S. Census Bureau, 2011). Technology is also advancing at a fast pace but is often designed without the elderly in mind and as a result has outpaced the ability of some in this population to understand and use it. In order to accommodate the elderly in the design process, the cognitive limitations of this population need to be understood and addressed. Improvements to technology have the potential to increase the quality of life of the elderly by helping them maintain independence and autonomy in our technology-driven world.

Response selection is concerned with the processes that are used in determining the appropriate response to a given stimulus. The effect of aging on response selection is one important cognitive limitation that is in need of greater research. The current study is intended to focus on a major gap in the aging and response selection literature; specifically how manual or vocal responses affect task performance of the elderly. The purpose of the current research is to determine the extent to which prior research on aging and response selection is applicable to real-world tasks. This would enable a better understanding of age-related deficits. In addition, this information could be used to aid in the design of technology that better accommodates the needs of the elderly and could be used as a guideline in future designs of products.

Response selection is important because it is a decision-making process that is involved in everyday life and applies to many domains such as driving and sports. In the simplest information-processing model, there are three information stages that are involved from the onset of a stimulus to a response, which are comprised of stimulus identification, response selection, and response execution (Proctor & Van Zandt, 1994). Stimulus identification is concerned with the perceptual processes involved in indentifying a stimulus through its features or patterns. After the stimulus has been identified, a person has to decide what response to make, which is referred to as the response selection stage. Last is the response execution stage; once a response has been selected, the selection is turned into a motor response.

The remainder of this section will be divided into two sub-sections. The first part will discuss significant prior research on aging and response selection. The second part will discuss the Interactive Voice Response system and the design needs of the elderly.

Aging and response selection in spatial choice tasks

In the elderly, response selection processes are the main cause of slowing when producing a response to a stimulus (Proctor et. al, 2005). Meiran and Gotler (2001) found that the duration of response selection processes caused age-related differences in task-switching performance. According to Sanders (1998) the duration of response selection is affected by four main factors: number of stimulus and response alternatives (S-R alternatives), stimulus-response compatibility, movement response precuing, and relative stimulus and response frequency (S-R frequency).

Number of S-R alternatives refers to the number of S-R pairings in a given task. For example, a person might be instructed to press one of three buttons, which are mapped to the locations left, right, and center, when they see a stimulus in a particular location on a screen. As the number of S-R alternatives increase, so does the reaction time required to respond, which is commonly referred to as the Hick-Hyman law (Hick, 1952; Hyman, 1953).

S-R compatibility (SRC) is defined as the degree to which a stimulus that is perceived is consistent with a response that needs to be made (Proctor, Vu, & Pick, 2005). An example of SRC can be found with the common door. If there is no knob or handle on a door, most people expect to be able push the door open. If there is a knob or handle on the door, the expectation is to turn the knob or grab the handle to open the door. In both cases, the stimulus and response are compatible; the stimulus elicits a person to make the correct response. In this example, stimulus response incompatibility would occur when a person encountered a door with a knob or handle that required a push, rather than a turn of a knob or grab of a handle to gain entry. The more consistent the stimulus is with the response, the better the performance.

Response precuing is when a subject is prepared for a response to an upcoming movement by being cued with a subset of possible responses in advance (Proctor et al., 2005). An example is a spatial precuing task, where subjects are precued with information about what finger from the left or right hand to use to respond to the forthcoming stimuli. In the task, a subject responds to spatial stimuli with a keypress that corresponds to the location of the stimuli. Keypress responses are made with one of four fingers, the index and middle finger from each hand.

Relative S-R frequency is defined as when the familiarity of a stimulus-response mapping results in a shorter response time. For example, when using an IVR to communicate with a company most people are familiar with pressing the “0” on a telephone keypad to speak to an operator. As a result, if callers were suddenly asked to press "1" to speak to an operator there would be a longer response time than if they were asked to press the more familiar "0" button. In addition, callers may make errors due to trying to press the more familiar button associated with speaking to an operator.

The most influential factors on the duration of response selection are SRC and movement response precuing (Sanders, 1998). As shown in Figure 1, SRC is comprised of multiple levels and components, some of which are beyond the scope of this paper. The current research will focus on spatial SRC effects, specifically the SRC proper effect.

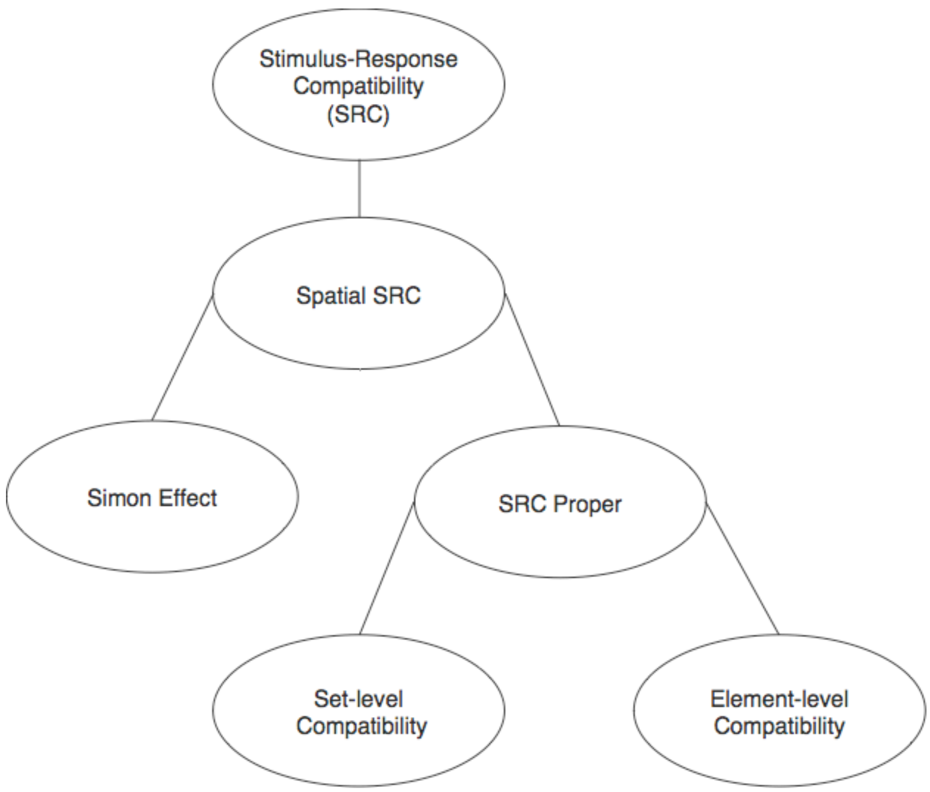


Figure 1. Stimulus-Response Compatibility.

Spatial SRC is concerned with the spatial relation between a stimulus and a response. The two types of spatial SRC effects are the Simon effect and stimulus response compatibility proper. The Simon effect occurs when the stimulus location is irrelevant to the task. In a simple Simon task, a stimulus (usually a color) is assigned to a particular response, a left or right button. Participants are asked to press the button corresponding to a stimulus that appears on the screen. An example of the task is shown in Figure 2, where participants are asked to press a left button when a blue stimulus appears on the screen. When the S-R are in the same relative location (blue stimulus on the left side of the screen), subjects respond more quickly. However, when the S-R are in opposite locations, subjects respond slowly. The findings on the Simon effect shows that subjects have a faster reaction time and are more accurate when the S-R are on the same corresponding side, even when the spatial position of the stimulus is irrelevant.

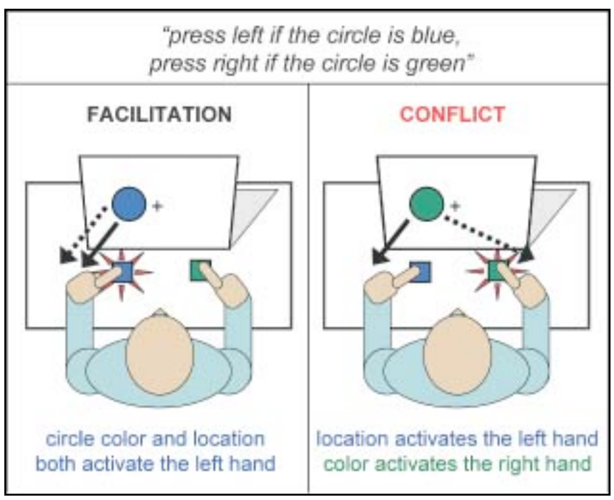
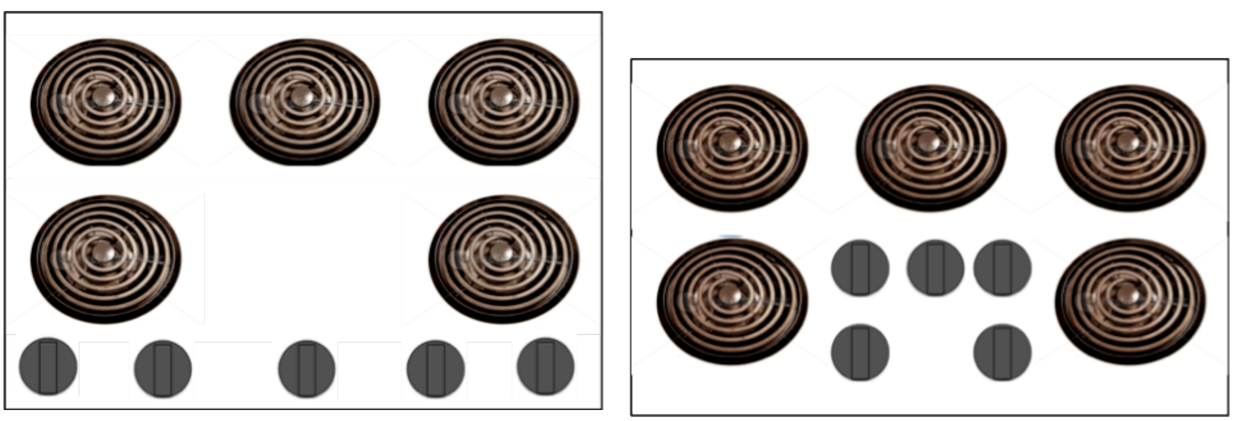


Figure 2. Simon Effect. Reprinted from "Neurcognitive mechanisms of action control: resisting the call of the sirens," by by K.R. Ridderinkhof, B. U. Forstmann, S. A. Wylie, B. Borís, W. Van den wildenberg, 2004, *WIREs Cognitive Science*, 2, p. 174.

SRC proper effects can occur when information about stimulus location is relevant to the task. According to Proctor et. al (2005), an example of the SRC proper effect is shown in Figure 3, which shows two images of stovetops with different burner and control layouts. When burners and controls are spatially arranged in the same manner, users are faster and more accurate at turning burners on and off. For example, users would respond faster to the stovetop shown in Figure 3b than Figure 3a due to the spatial correspondence between the burners and controls. Research conducted by Chapanis and Lindenham (1959) aligns with this example; as they found that users are slower at using stovetops with linear controls as shown in figure 3a (as cited in Proctor et. al, 2005, pg. 253).

The SRC proper effect is obtained when there is a faster and more accurate response to a stimulus due to the spatial correspondence between a S-R pairing (Hommel, 1995). The SRC proper effect is of most concern to people in the field of human factors because the relation between display and controls can affect the usability of products (Kantowitz, Triggs, & Barnes, 1990). Many designers have included the rules of SRC proper into design guidelines for better performance (Vu & Proctor, 2001).



(a) (b)

Figure 3. An example of the SRC proper effect using an oven stove top. (a) Five burner stove top with a linear arrangement of controls; (b) five burner stove top with controls spatially arranged in the same manner as the burners.

There are two types of SRC proper effects, element- and set-level compatibility (Kornblum, Hasbroucq, & Osman, 1990). The element-level compatibility effect occurs when alternative mappings between a single stimulus set and response set causes a difference in response time (Wang & Proctor, 1996; Proctor & Wang, 1997). In a single S-R set, a congruent mapping would be considered to have high element-level compatibility, whereas an incongruent mapping would have low-element compatibility. For example, given a two-choice reaction task, where a person has to press a button that is located on the left or right when they see stimuli on the left or right side of a screen, a task that requires a left button press to a left stimuli would be considered a congruent mapping, whereas a task that requires a left button press to a stimuli located on the right side of the screen would be considered incongruent. Wang and Proctor (1996) showed that on a two-choice reaction task subjects had greater accuracy and a faster response time with a congruent stimulus-response mapping than with an incongruent mapping.

Set-level compatibility is defined as a difference in response time that occurs when different combinations of high element-level compatibility S-R sets are compared to one another. Set-level compatibility is best described using the concept of dimensional overlap, which is defined as "the degree to which two sets of items have properties or attributes in common, and the degree to which such attributes are similar to one another" (Kornblum, 1992 p. 749). S-R sets that have more attributes in common would be considered to have high set-compatibility. For example, high set-compatibility would occur when the stimulus is auditory and the response is verbal, the commonality being the sound. Conversely, low set-compatibility would occur when the presented mode of the stimulus is auditory and the response is manual. Overall, dimensional overlap promotes more rapid response selection (Kornblum, Hasbroucq, & Osman, 1990). When a given S-R set has dimensional overlap, an automatic response is activated. Specifically, when a stimulus is presented, direct activation from long-term memory S-R associations automatically produces a response. An automatic response is fastest when the response in long-term memory matches the assigned response and is slower, due to interference, when there is no match (Proctor et al., 2005).

The current research is concerned with examining the effect of aging on SRC proper and response modality (verbal and manual). Most studies that examined this concept have lacked ecological validity; the tasks that have been tested have little relevance to the real world. In addition, different tasks yield slightly different results as will be discussed later. As such, it is unclear whether the results from this type of research can be generalized to the real world, across different tasks and age groups. The current research will seek to examine these issues. The remainder of this section will discuss related studies that have been conducted using different age groups. The age groups used were younger (18 to 39 years old) and older adults (65 years and older).

Doose and Feyereisen (2001) found a significant interaction between age (younger and older subjects) and response modality (verbal and manual) on a conceptual comparison task. In the task, subjects had to compare two pictures or words that were presented on a computer screen in succession. Subjects were instructed to determine if the presented pictures or words had the same meaning. In the task, a verbal and manual response was required. Subjects verbally responded by saying yes/no or manually responded by pressing a button. Younger subjects responded faster with a manual response (i.e. low set-level compatibility), while older subjects were faster with a verbal response (i.e. high set-level compatibility). Error rates for both groups were about the same.

Emery (2006) conducted a study in which older and younger subjects viewed a sequence of letters or locations and had to recall what they had seen via a verbal or manual response. Older subjects had a higher verbal span on the sequence of letters task with a verbal response (i.e., high set-level compatibility) than with a manual response (i.e. low set-level compatibility). Recall on the location task was the same when the response modality was verbal or manual for older adults (i.e., no effect of set-level compatibility). Younger adults had equivalent spans regardless of response modality (i.e., no effect of set-level compatibility).

Overall, the results from Doose and Feyereisen (2001), Wang and Proctor (1996), and Emery (2006) seem to conflict; each study obtained different results for younger adults on set-level compatibility tasks. The results obtained for older adults were the same for Doose and Feyereisen (2001) and Emery (2006), who compared younger and older adults on set-level compatibility tasks. These studies found that older subjects performed better on high set-level compatibility tasks with a verbal response. In addition, the elderly performed worse on low set-level incompatibility tasks with a manual response. The difference in results could also be due to task differences, which is of interest to the current research because of how set-level compatibility results may vary across tasks.

In keeping with the goal of studying SRC proper effects within the context of a more real-world task, an Interactive Voice Response (IVR) will be designed for use in the current research.

Interactive Voice Response Systems

The technology of interest is the Interactive Voice Response (IVR) system, which is commonly used to complete everyday tasks, such as banking, checking the weather forecast, and paying utility bills. The IVR is a technology that allows people to use a telephone to obtain information from a computer database. An IVR is used to greet and communicate with customers in many industries such as banking, healthcare, and travel. The key responsibilities of a typical IVR system are to take a call, ask a series of questions to direct the caller to the correct option, and handle the caller's voice and/or keypad inputs. Callers can interact with the IVR through a speech-enabled or touch-tone interface.

A benefit provided to the customer through the IVR is increased privacy when handling personal matters over the phone. The advantage of an IVR is often outweighed by a poor design. A poorly designed interface causes frustration and confusion for everyone, especially for the elderly whose limited cognitive abilities make interaction and recovery from error even harder. Variance due to a poorly designed IVR must be eliminated in order to properly conduct the current research. As a result, the remainder of this section will focus on cognitive abilities required to use the IVR and design guidelines that will be followed to address these issues.

IVR systems are often designed for the general population rather than a specific group (Brandt, 2008). As a result, many of these systems fail to account for the potential cognitive limitations of specific customer segments, such as the elderly. According to Dulude (2000), in order to use an IVR one must have sufficient memory and attentional resources, as well as the ability to process information quickly (speed of information processing). A person's working memory capacity can affect the ability to successfully interact with the IVR. As people age, performance declines on working memory tasks that require active processing along with passive storage (Verhaeghen, et al. 1993). In addition, older adults perform worse than younger adults on complex tasks that place more demands on working memory (Salthouse, Mitchell, Skovronek, & Babock, 1989). Using an IVR would be considered a working memory task, since users are required to process, store, and manipulate a large amount of information in a short period of time. When interacting with an IVR, a caller is provided with a series of choices for which they must provide a response. For example, a typical bank IVR provides a caller with choices such as, "press 1 for checking" and "press 2 for savings." A caller must use working memory to hold each of these options and corresponding responses (i.e., the stimulus-response set) active in memory in order to make a correct choice. Response modality may also affect the caller's ability to hold and recall information, depending on the task and the user's memory capacity as discussed in Emery (2006).

According to Milham et. al (2002), working memory issues in the elderly are the result of the brain's inability to control attention. Lack of attentional control can cause vulnerability to the distraction of irrelevant stimuli. As a result, people who cannot selectively control attention may not be able to attend to important information that may need to be recalled and manipulated later. Use of an IVR requires attentional control in that a caller must be able to filter out distractions occurring in the background, as well as hear prompts to navigate the system to their desired option. Kline and Scialfa (1997) have shown that maintaining auditory information is harder for older adults than younger adults.

Speed of processing refers to how rapidly someone can process incoming information and execute a response; it declines with age. Salthouse (2000) showed that older adults are slower at performing speeded tasks due to age-related changes in processing speed. In a sense the IVR is a type of speeded task in that a prompt is presented and a caller has a limited time to process the information and make a correct choice.

Another important cognitive ability required for IVR use is spatial ability, which refers to the cognitive processes that are involved in performing actions, such as perceiving, recalling and transforming a visual image (Lohman, 1996). An example of a spatial ability task is the paper-folding test as shown in Figure 4. In this test, subjects are shown a sequence of folds in a piece of paper. In each fold is hole that was punched through the folds of the paper. The subject has three minutes to determine from a set of unfolded papers, which holes corresponds to those seen previously.

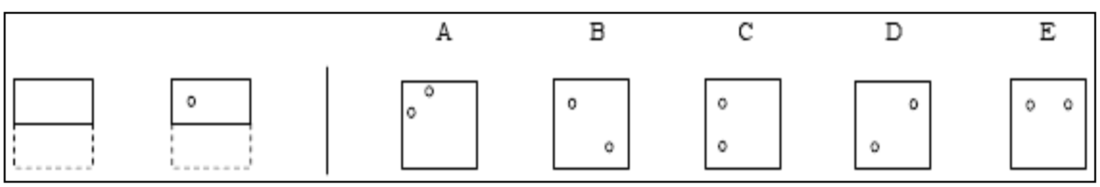


Figure 4. Paper Folding Test. Reprinted from "Kit of Factor-Referenced Cognitive Tests," R. B. Ekstrom, J. W. French, H. H. Harman, and D. Derman, 1976, Princeton, NJ: Educational Testing Service.

The relationship between spatial ability and performance has been observed in a wide range of visual display tasks such as searching for information on the internet and using a computer to edit documents (Pak, Rogers, & Fisk, 2006; Sharit, Hernandez, Cjaza, & Pirolli, 2008). Prior research has shown that correct mental models used on search tasks can be helpful in successfully finding information (Borgman, 1986). However, creating mental models may place substantial demands on spatial ability, which has been shown to decline with age (Salthouse, 1992). Pak, Czaja, Sharit, and Rogers (2008) discovered that a decline in spatial ability accounted for performance differences on an auditory task. In the study, they had older and younger adults perform basic tasks on an IVR system and found that the variance in performance was partially due to age-related differences in spatial ability between groups.

Several interesting hypotheses arise out of the literature review on the effect of aging on response selection and this research will seek to examine these with the aid of a customized IVR system that will take into consideration the cognitive limitations of the elderly population.

Hypotheses

H1 The difference between high set and low set-level compatibility is larger for older than younger subjects in an IVR task.

S-R sets that have more attributes in common would be considered to have high set-level compatibility. For instance, a task with a auditory stimulus-vocal response would be considered to have high set-level compatibility, while low set-level compatibility involves the use of a auditory stimulus-manual response. Previous findings show that performance on set-level compatibility tasks for younger adults tend to vary, while older adults consistently perform better on tasks with high set-level compatibility than low set-level compatibility. Since the current research will use an IVR, which involves set-level compatibility because subjects can interact with the system through a speech-enabled (auditory stimulus-verbal response) or touch-tone interface (auditory stimulus-manual response), I surmise the results will generalize to tasks performed on the IVR.

H2 *The difference between mapping type (indirect vs. direct) will be larger for older adults than younger adults.*

When using an IVR a caller usually has the option to verbally or manually respond to an auditory prompt. On certain tasks, one response modality may lead to better performance than the other depending on the mapping required. Typically mapping is required when using a manual response, where a value is mapped to a telephone keypad. One hypothesis is that regardless of response modality, mapping is not involved when a value is input (e.g., highly used number, such as a telephone number or social security number) that has been committed to long-term memory. The person is simply saying or typing their well-memorized number from memory onto a telephone keypad. In the current research, this mapping type is referred to as direct. However, it is assumed that mapping is required when a response has not been committed to memory and/or the user has to indirectly input the value. For instance, in a given IVR system a person may be instructed to input their prescription number, which is infrequently used. As a result, the prescription number is read from a statement or bottle and then inputted via a verbal or manual response. This mapping type would be called indirect in the current research. If the hypothesis is correct an interaction between age, response modality and mapping type will be found. Mapping should negatively affect older adults performance more than younger adults on tasks.

METHOD

Subjects

A total of 50 subjects participated in the study: 25 older adults (65 years and older) with a mean age of 71.7 (*SD* = 4.9) and 25 younger adults (18 to 39 years old) with a mean age of 27.3 (*SD* = 5.2). Subjects in the older adult group consisted of 7 males and 18 females. In the younger group there were 8 males and 17 females. Subjects were diverse in terms of education and ethnicity. Forty-nine out of fifty subjects had some type of higher education, which ranged from some college to a doctorate degree. The ethnic background of the subjects was as follows: 64% Caucasian (*n* =32), 24% African American (*n* =12), 4% Asian (*n* =2), 4% Hispanic (*n* =2) and 4% Biracial (*n* =2). All subjects had normal or corrected-to-normal vision and were fluent English speakers. The majority of subjects, 98% (*n = 48*), reported using an IVR system in the past and 68% (*n* =32) reported using an IVR at least once in the past month.

Subjects were recruited from the larger Houston population and were primarily recruited from a paid advertisement on Craigslist online. An additional recruiting method included contacting people from the Rice University Psychology subject pool. Subjects were compensated $25 cash for their participation in the 75-minute study.

Design

A mixed design with one between-subjects variable and three within-subjects variables was used. The between-subjects variable was age. The three within-subjects variables used were task (healthcare and banking), response modality (verbal and manual), and mapping type (direct vs. indirect). Each prompt was delivered auditorily from the IVR system, and the response modalities used were manual and verbal. Mapping type is based on the hypothesis that regardless of response modality, mapping is not involved when inputting a number that has been memorized. For this study, two mapping types were used: direct and indirect. It is assumed that direct mapping is not involved when an input value (e.g., a highly used number, such as a telephone number or social security number) has been committed to long-term memory. The person is simply speaking or typing their well-memorized number onto a telephone keypad. Indirect assumes that mapping is required when a response has not been committed to memory and/or the user has to indirectly input a value. For instance, in a given IVR system, a person may be instructed to input their prescription number, which is infrequently used. As a result, the prescription number is read from a statement or bottle and then inputted via a verbal or manual response. Using a combination of the response modalities (i.e., verbal and manual) with mapping type there were a total of four conditions: manual response-direct condition, manual response-indirect condition, verbal response-indirect condition, and verbal response-direct condition. Each condition was completed for both the healthcare and banking domains as shown below in Figure 5.

Tasks (2)	X	Response Modality (2)	X	Mapping Type (2)
• Healthcare • Banking		• Manual • Verbal		• Direct • Indirect

Figure 5. Methodology design of the experiment showing the within-subject variables.

In the study, there were two main performance metrics: response time and error count. Response time is the duration taken to complete a section of the IVR. Response time was measured in seconds and automatically recorded by the IVR application. In the IVR, time stamps were recorded at the start and end of a call and after each decision point. Error count is the sum of errors that could occur in a particular section of the IVR. The types of errors are shown in Table 1 and 2. The sections of the IVR are balance, navigation, and total. Balance is the section of the IVR where the account balance is checked and involves entering a 4 and 10-digit number (this will be discussed in detail in a later section) to hear the balance. Navigation refers to the part of the IVR where the subject navigates the IVR until the subject checks their balance. Total is the duration of the whole call, which equals navigation plus balance. Total was originally defined as the duration from the start of the call to the end. However, a majority of subjects hung up after hearing their balance, which occurred before the last time stamp was recorded. Therefore, total was redefined as the start of the call until the balance was heard (see Appendices D – G). The redefinition of total reflects what commonly occurs in the real world, since most people hang up after hearing their desired information.

Table 1

Input Error Types for checking balance in the IVR system

4 and 10-digit input errors
• Manually inputted code wrong (system) • Verbally inputted code wrong (system) • No entry (system) • Timed out • Wrongfully entered own birth date • Wrongfully entered own phone number • Manually inputted code instead of saying • Verbally inputted code instead of manual

Table 2

IVR Error Types for navigating the IVR system

IVR menu errors
• No entry (system) • Invalid entry (system) • Manually inputted option instead of saying • Verbally inputted option instead of manual • Timed out • Kicked out (system)

All subjects completed three tests to assess whether or not they possessed the cognitive abilities needed to use an IVR. The cognitive ability tests were completed in the following order:

- Spatial Ability

  The Paper Folding test (Ekstrom et. al 1976) measures spatial ability. In this paper and pencil test, 20 problems are given to subjects. Each problem shows an image of a sequence of folds in a piece of paper. In each fold is a hole that was punched through the folds of the paper. The subject has a total of six minutes to determine from a set of unfolded papers, which holes corresponds to the one they have just seen. The score for this test is calculated by subtracting the number of items correct from the number incorrect.

- Working Memory

  The Digit Span (WAIS-III; Wechsler, 1997) is a subtest in the working memory index in the Wechsler Adult Intelligence Scale (WAIS). In this test, subjects are read a series of digits and then are required to orally recall the digits forward and backwards. The test is not timed. The test is scored by adding up the points from the digits forward (16 points) and backward (14 points) tests to get a total digit span score (30 points).

- Working Memory and Attention

  The Letter-Number Sequencing (WAIS-III; Wechsler, 1997) is a subtest in the working memory index in the Wechsler Adult Intelligence Scale (WAIS). This test measures attention and short-term auditory memory. In the test, subjects are read a series of numbers and letters in random order. The subject must then orally recall the numbers first in ascending order and then the letters in alphabetical order. Giving 1 point for every correct trial scores the test. There are a total of 21 points that can be earned. This test does not have any time restrictions.

Task

As shown in Figure *5*, subjects completed four-response modality (verbal and manual) by mapping (direct and indirect) conditions for each task domain (banking and healthcare). In each condition, subjects called into an IVR and navigated the system to check an account balance using a ten and four digit account number. Each participant made a total of eight phone calls to an IVR system. Instructions for each call were written on index cards that corresponded to each condition. The instruction cards were given to the subjects one at a time and were counterbalanced by condition. Each instruction card contained the following:

- A description of the scenario (a story to setup the task domain)

- The phone number of the IVR system to call
- The response modality to be used for input
- Mapping type for the ten and four digit account numbers
- End of condition instructions

The scenario for each instruction card differed (see Appendix A), but involved the same central task of checking a balance. In the banking domain, subjects checked a checking account balance. In the healthcare domain, subjects checked a balance owed for hospital services rendered. The mapping types used were indirect or direct (discussed in detail in the design section). In the direct conditions, subjects were instructed to use their personal ten-digit telephone number (area code and phone number) and four-digit birthday (month and day) as account numbers. In the indirect conditions, subjects were instructed to use the ten and four digit number written on the card as account numbers. After the completion of each condition, subjects were instructed to write the balance they obtained from the IVR on an answer sheet.

IVR Design

The IVR was designed to accommodate age-related cognitive ability changes. In general, common problems experienced with IVR systems are poor menu organization, lengthy instructions, difficult to understand voice prompts, and problematic error recovery. Reducing these common problems combined with techniques that take cognitive limitations into account should yield an IVR that can more efficiently and effectively be used by older adults. The first issues addressed were the difficulty in understanding voice prompts and lengthy instructions. Generally in an IVR, there are two types of voices that can be used to present the prompts to the user, synthetic and prerecorded natural human speech. Synthetic speech can vary in sound from a computerized to naturalistic human voice and is produced by a computer (Evans, 2009). A naturalistic male synthetic human voice was used to deliver the prompts in the current IVR experiment. A male voice was used for the current study; previous research shows that voice gender does not affect IVR usage (Evans, 2009)

Another problem addressed in the IVR design was menu navigation. In IVR systems, there are two types of menu structures: deep and broad. In a deep menu structure, only a few or more menu options are placed on a level, which causes the menu structure to be more than four levels deep. In a broad menu structure, all menu options are placed on one or more levels. Both deep and broad menu structures make it difficult for customers to remember menu options, and as a result time is wasted replaying menu options or listening to a prompt irrelevant to the user's task. Irrelevant menu options make it necessary for users to utilize more attentional resources. This can be especially challenging to the elderly who tend to have fewer attentional resources. Prior research shows that only four menu options at a time should be presented to a caller and that each menu option should be no deeper than three levels (Engelbeck & Roberts, 1990; Gardner-Bonneau, 1999). A recent study by Commarford, Lewis, Smither and Gentzler (2008) reexamined the Engelbeck et. al (1999) IVR guideline. In the study, a deep menu structure was compared to a broad menu structure using a speech-enabled interface. The results showed better performance and a higher satisfaction with a broad menu structure. Commarford et. al (2008) suggested participants do not store all of the menu options in their memory as previously thought, but listen to the prompts and only store the menu options that are the most relevant to their desired goal. The menu structure that enables subjects to eliminate irrelevant menu options faster and easier is the broad menu structure. Commarford et. al (2008) also found that broad menu structures are better for individuals with lower working memory capacities. While the Commarford et. al (2008) research did not involve a manual IVR interface, previous research provides evidence that broad menu structures are better than deep menu structures for manual response IVR systems as well (Huguenard, Lerch, Junker, Patz & Kass, 1997; Virzi & Huitema, 1997). The Commarford results are in contrast to a recent study by Charness and Jastrzembski (2009), who found that auditory menus should use a deep menu structure rather than a broad structure, since menu options are unknown in advance. For the current study, the research by Engelbeck et. al (1999) and Commarford et. al (2008) were used as an IVR guideline. A broad menu structure that was no more than three levels deep was implemented in the current research. In addition, menu options were presented to the user by importance or popularity. The importance or popularity of a menu option depends on the most common transactions for a particular domain. For instance, a person might call a banking IVR the most for checking the balance on their savings account, while for a hospital they might commonly call for patient billing. Both menu options would be displayed on the top level.

Materials and procedure

IVR System

QuickFuse (www.quickfuseapps.com), a web-based voice application and editor created by Plum Voice, was used to create IVRs for the study. This commercial telephony application, which utilizes Dual Tone Multi-Frequency (DTMF), speech recognition, and text-to-speech technology, allows a full scale IVR to be built easily with a drag and drop editor. Also, QuickFuse includes a cloud database that can be used in conjunction with IVR applications that are created.

IVR System Design

For the study, IVRs were created for each task domain and response modality (i.e. banking-verbal, banking-manual, healthcare-verbal, healthcare-manual – see Appendices D - G). In the verbal conditions, subjects were instructed to speak their response. In the manual conditions, the IVR instructed subjects to use the telephone keypad to enter their response. All the IVRs created were the same except for the content of the prompts that reflected the different task domains and response modalities. In the IVRs, there were three major decision points: the main menu, the account menu and the end call menu. A decision point is defined as point in the IVR in which a decision has to be made.

At the start of the call, a time stamp was recorded and the main menu was presented to the subject. This top-level menu had eight menu options, and only one of the menu options allowed subjects to advance in the IVR. The menu option that allowed advancement was positioned in the same place in all IVRs created for the study (menu option 6). The menu options that did not allow advancement were maintenance nodes; when the subject selected one of these menu options, the IVR prompt would inform the subject that the system is currently being updated and return the subject back to the main menu. The addition of maintenance nodes (non-advancement menu options) gave the illusion of a full scale IVR system. If a single maintenance node was selected more than three times, the IVR system would terminate the call. The main menu did not have a barge-through option, and the subject would have to listen to all of the menu options before making a selection. After all the menu options were heard, a secondary prompt with a barge-through option instructed the subject to enter a response. The main menu had two system errors associated with selecting a menu option, silence (no entry) and invalid entry. The no entry error occurred when a subject failed to make a selection. The invalid entry error occurred when a subject made an invalid selection. If a single error occurred more than three times, the IVR system would terminate the call.

After the subject selected the correct menu option, they were advanced to the account menu (the name of the account differed for the banking and healthcare task). The account menu had two menu options: a maintenance node and option that allowed the subject to check their account balance (balance menu). The maintenance node informed the subject that the system was currently being updated and returned the subject back to the main menu. The account menu did not have a barge-through option, and the subject would have to listen to all of the menu options before making a selection. After all the menu options were heard, another prompt with a barge-through option would instruct the subject to enter a response. The account menu had two system errors associated with selecting a menu option: silence (no entry) and invalid entry. When the balance menu option was selected, a time stamp was recorded. Then, subjects were instructed to enter a ten digit account number and subsequently a four digit account number. The account numbers were queried from the database associated with the IVR. Each account number input point had two system errors associated with it: invalid entry and unable to find. The invalid entry error, in this case, referred to unexpected events that may be encountered, such as a table no longer existing or an invalid query parameter. The “unable to find” error occurred when no rows matched the conditions specified in the query made. If a single error occurred more than three times, the IVR system would terminate the call. After the account numbers were input successfully, the account balance was heard and a time stamp was recorded. The last decision point in the IVR was the end call menu. This menu had two options: repeat balance and terminate the call. The repeat balance menu option allowed the subject to hear the account balance again. The termination menu option allowed subjects to end the call. The end call menu did not have a barge-through option, and the subject would have to listen to all of the menu options before making a selection. After all the menu options were heard, another prompt with a barge-through option would instruct the subject to enter a response. When subjects chose to terminate the call, a time stamp was recorded and all of the data collected from the call was sent to a database.

IVR Workarounds

The module used for the secondary main menu did not allow for a specific response modality to be designated. As a result, subjects could enter a response verbally or manually regardless of the response modality condition. This problem was circumvented by covering the telephone keypad with a piece of paper labeled “DO NOT USE KEYPAD” during the verbal condition. The telephone keypad cover was primarily used when the subject attempted to enter a response manually during a verbal condition. During the manual conditions, the experimenter observed the subject and intervened if the subject attempted to enter a verbal response.

In the QuickFuse database, leading zeros were removed. As a result, subjects whose birth month began with a zero, as in the month of April, were instructed to input their birth month as “14” rather than “04.” Subjects were reminded to replace the zero with a one in the PowerPoint instruction slides and on the instruction index card.

Experiment Setup

The experiment setup involved a video camera, dry erase board, computer and phone as shown in Figure 6. The video camera was used to record the subject's hand and audio information while completing the experiment. On the date of the session, the subject's identification number and condition were written on the dry erase board in view of the video camera. A computer was used by the experimenter to read instructions from slides to guide subjects through the experiment and to record notes about the session. The phone, a corded AT&T CL4939 with extra large buttons, large display, and receiver volume control, was used by the subjects to call into the IVR as shown in Figure 7. During the experiment, the receiver volume of the phone was set on the highest setting, which enabled the video camera to pick up sound and the experimenter to better moderate the task.

Figure 6. Setup of experiment.

Figure 7. AT&T CL4939 phone used in the experiment.

Procedure

Prior to the start of the experiment, the subject's four-digit birthday (month and day) and seven-digit phone number was entered into the QuickFuse database.

The study took place in a single 75-minute session. Upon entering the experiment room, subjects were seated in front of the phone and were read instructions from a presentation (see Appendix C). The instructions ran through each part of the experiment. The experiment was divided into three parts: cognitive ability tests, condition completion, and debriefing. The first section of the experiment involved subjects completing three cognitive ability tests mentioned earlier in the text. Following the completion of the cognitive ability tests, subjects were read general instructions on how to complete the conditions in the experiment. After the subject was given an opportunity to ask questions, the experimenter started recording with the video camera. Then the second part of the experiment, completion of the trials, began. At the beginning of each trial, subjects were handed an instruction card. After reading the instruction card and asking any questions, subjects dialed the number on the card. At the end of each trial, subjects were instructed to write the information they received from the IVR on an answer sheet (see Appendix B). Once all of the trials were completed, subjects were instructed to fill out a questionnaire and were debriefed. The questionnaire collected demographic information prior IVR usage, the opinion of the IVR used, the task, and overall experience in the study.

RESULTS

A 2 (age group) x 2 (tasks) x 2 (response modality) x 2 (mapping type) x 3 (education) ANOVA was performed on each IVR performance metric. The performance metrics were response time and error count, which measured particular sections in the IVR: navigating the IVR system (navigation), checking an account balance (balance) and total call. Seven subjects were excluded from analyses for the following reasons: failure to follow instruction (n = 3), call terminated early in two or more conditions (n = 3), or they were an outlier on more than three conditions for a performance metric (detailed outlier definition below) (n = 1).

Response Time

An outlier method designed to handle response time data was used for the response time variables. The subject mean across conditions (eight), was computed for each variable. Then, subtracting the subject mean from each observation, across conditions, created eight new variables. Outliers, defined as observations falling outside three interquartile ranges (IQRs) from the 25th or 75th percentiles of the distribution of the response time variables, were identified from the box plots. Subjects with three a more outliers on a response time variable were excluded from analyses. Observations were temporarily removed from the original data for subjects who had two or less outliers. Then, the subject mean was recomputed (without the outliers) and then used to replace removed observations. For the total time variable, 19 observations were replaced. A total of 18 observations were replaced for the balance time variable. The navigation variable is subtraction of the balance time from the total time variable, so no extra outliers were removed for this variable.

The response time variables were adjusted to accommodate a response delay in the IVR system. Time data from ten randomly selected subjects was examined to determine if there was an IVR response delay for verbal and manual responses. The IVR system takes on average, two seconds to respond to a manual response and six seconds for a verbal response. Since, the lag time differed for each response modality and the response time variables were to only reflect cognitive and motor time, the response delays were subtracted from the response time variables (see Appendix I).

A 2 (age group) x 2 (tasks) x 2 (response modality) x 2 (mapping type) x 3 (education) x 2 (replacement type) ANOVA was performed on the balance time variable. The replacement type variable refers to whether or not subjects had to replace the first digit in their month of birth with a one. This replacement took place while the subject was checking their account balance. In the study, a total of 32 subjects replaced the leading zero of their birth month with a one.

Errors

The error variable used for analyses is count variable, which is the sum of all errors that occurred in a particular section of the IVR. A square root transformation was used on the error variable, before analyses were performed. Two observations were excluded from these analyses because of missing data. Outliers for these variables were not removed, because over half of the subjects did not make any errors.

Hypotheses

The hypotheses for this study made predictions about interactions that would occur while navigating the IVR system (navigation) and checking an account balance (balance): response modality by age and response modality by age by mapping type. These sections of the IVR were measured with response time and error count.

Navigation

The interaction predicted for navigating the IVR system was response modality by age; it was expected that older subjects would perform worse in the verbal prompt-manual response (low set-level compatibility) condition than younger subjects. As shown in Figure 8 and 9 the prediction was not supported, performance for older and younger adults was approximately the same across response modality. The interaction between response modality and age, $F(1, 37) = 1.05$, $p = .31$, MSE = 21.89, $\eta p^2 = .03$ for the response time analysis was found to be statistically non-significant. No such interaction was found for errors when navigating the IVR also, $F(1, 37) = .01$, $p = .94$, MSE = .06, $\eta p^2 = .0$.

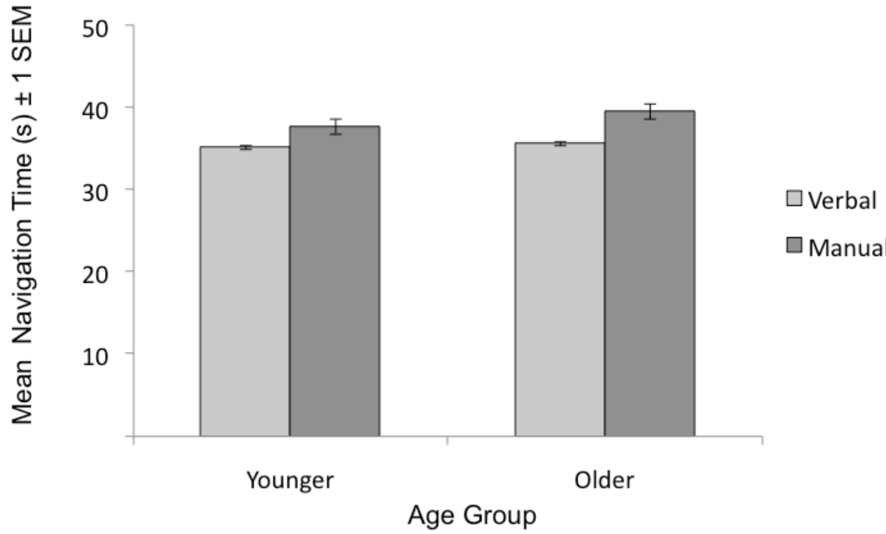


Figure 8. Mean navigation time for response modality by age.

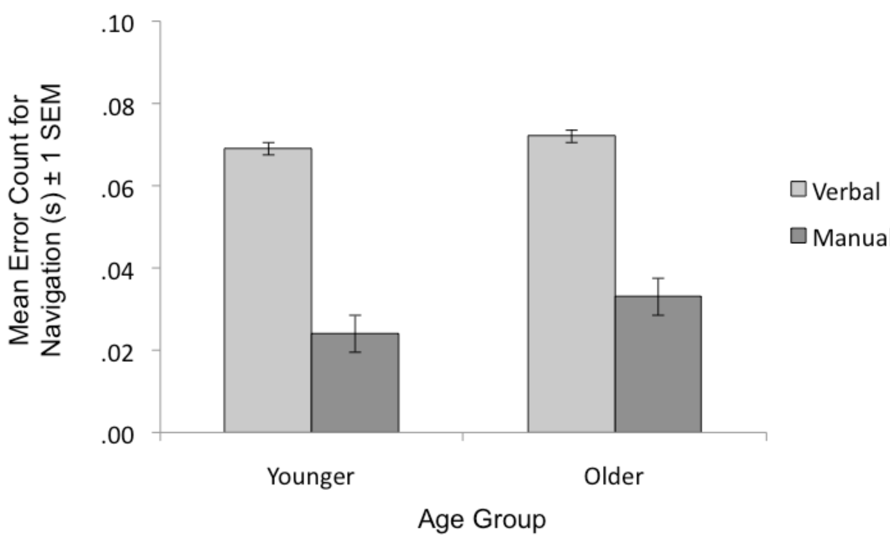


Figure 9. Mean error count for navigating the IVR system for response modality by age.

Balance

The interaction predicted for checking an account balance in the IVR system was response modality by age by mapping type; it was expected that mapping type would negatively affect older subjects performance more than younger subjects. As shown in Figure 10 – 11 and Figure 12 – 13 the prediction was not supported, there was only a slight difference between performance for older and younger adults. The interaction between response modality, age, and mapping type $F(1, 33) = .13$, $p = .72$, MSE = 10.82, $\eta p^2 = .00$ for the response time analysis was found to be statistically non-significant. Also, no such interaction was found for errors, $F(1, 32) = .68$, $p = .41$, MSE = .10, $\eta p^2 = .02$ However, a two-way interaction was found between response modality and mapping type, $F(1, 33) = 8.98$, $p < .01$, MSE = 8.36, $\eta p^2 = .27$. Figure 14 shows a slight difference in mapping type for the verbal condition, and as expected the cost of mapping type was larger for the manual condition. The difference between the verbal and manual condition was expected, since it takes longer for a person to input a value manually than vocally. This result aligns with the simple effect post-hoc test, which shows an effect of mapping for the manual response modality $F(1, 33) = 32.76$, $p < .01$, MSE = 4.48, $\eta p^2 = .48$ but not for the verbal response modality $F(1, 33) = .00$, $p = .96$, MSE = 10.55, $\eta p^2 = .00$. Given the nature of this interaction, it is not surprising that the main effects of response modality $F(1, 33) = 63.06$, $p < .01$, MSE = 10.69, $\eta p^2 = .67$ and mapping type $F(1, 33) = 10.75$, $p < .01$, MSE = 6.67, $\eta p^2 = .24$. are statistically significant.

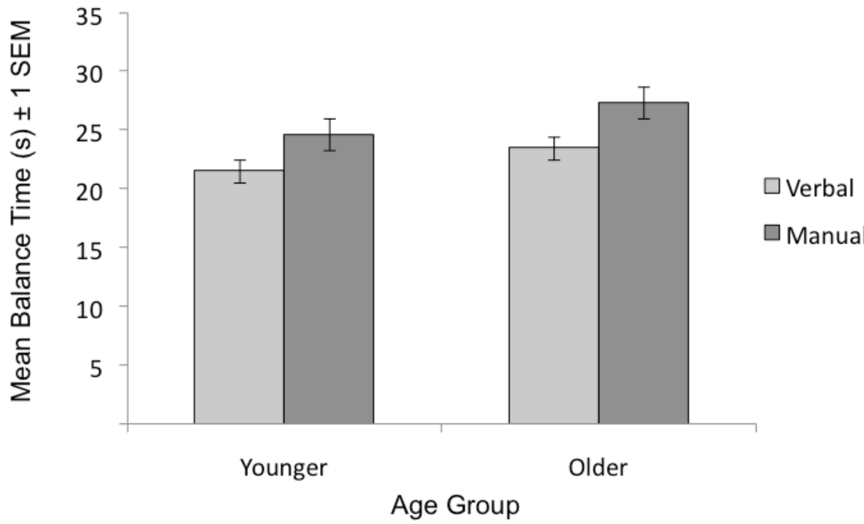


Figure 10. Mean balance time for age, response modality for the direct mapping type.

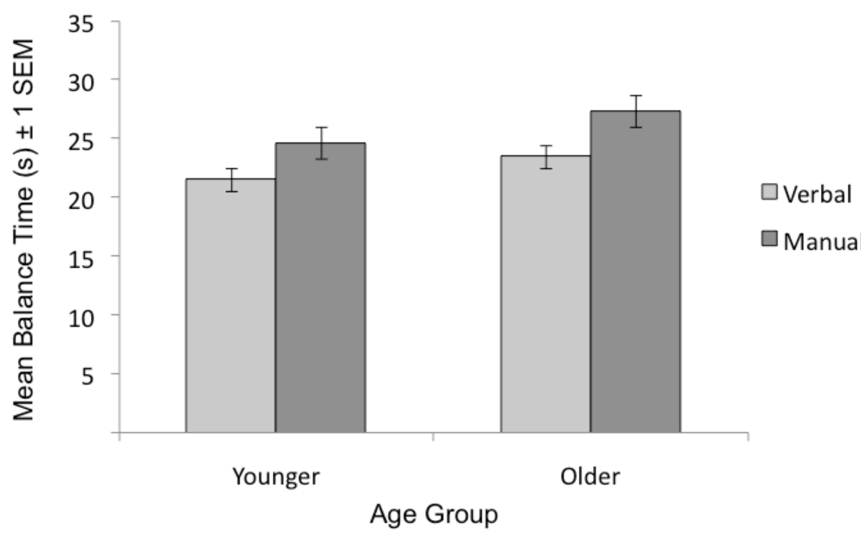


Figure 11. Mean balance time for age, response modality for the indirect mapping type.

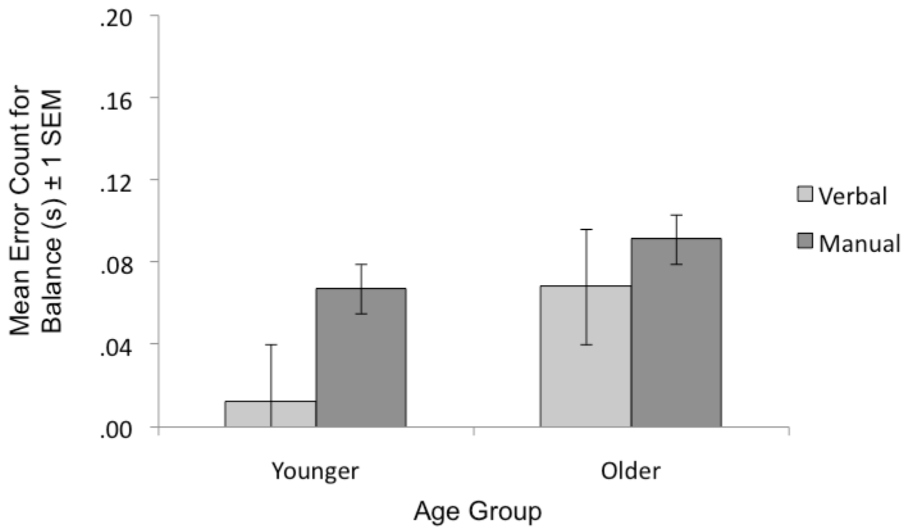


Figure 12. Mean error count for balance time for age, response modality for the direct mapping type.

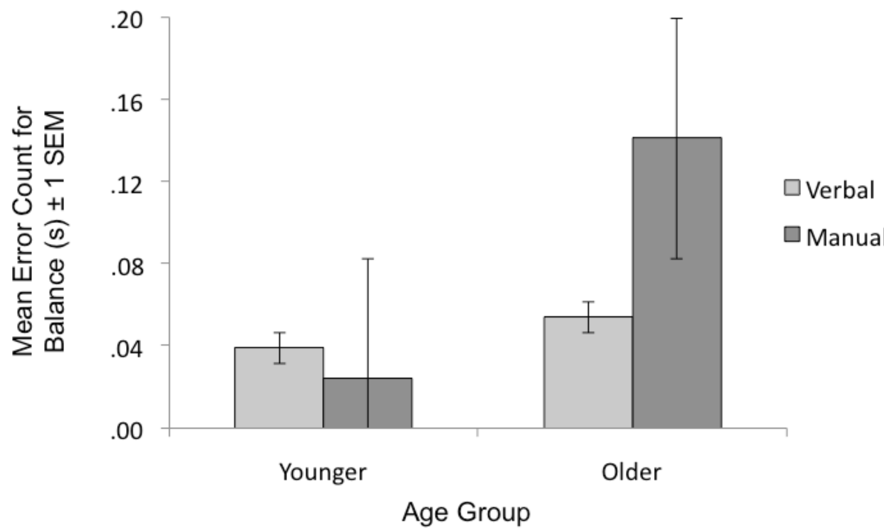


Figure 13. Mean error count for balance time for age, response modality for the indirect mapping type.

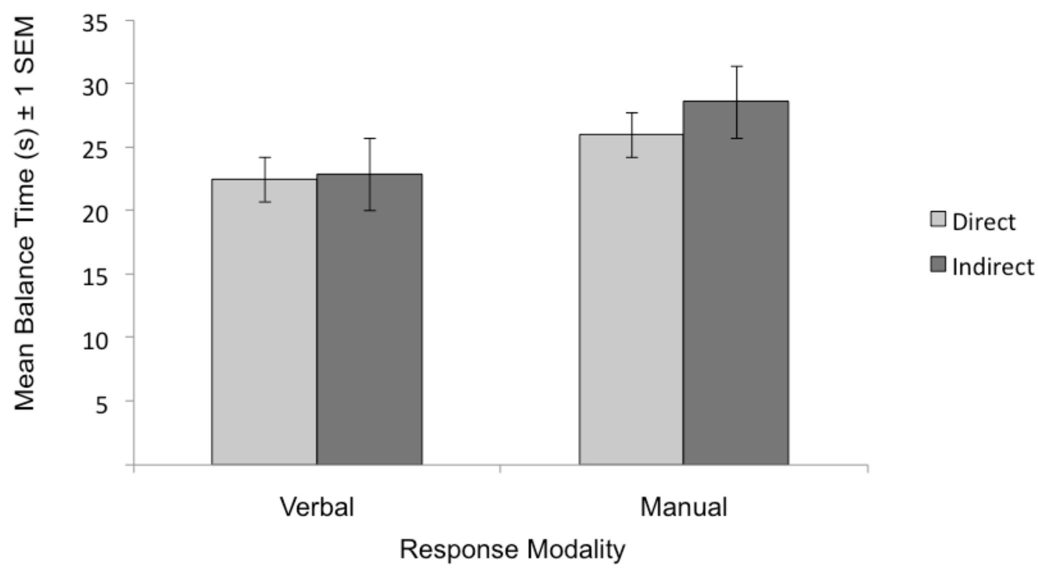


Figure 14. Interaction between response modality and mapping type.

Other results

Other results unrelated to the central hypotheses were obtained with the response time and error count variables corresponding to navigating the IVR system (navigation), checking an account balance (balance) and the total call (Navigation and checking an account balance).

Navigation - Response Time

Task

The mean times to navigate the banking and healthcare IVR were 38.83 (SD = 6.09) and 35.83 (SD = 5.42). The banking IVR took slightly longer to navigate than the healthcare IVR. The main effect of task $F(1, 37) = 15.92$, $p < .01$, MSE = 24.78, ηp^2 = .30 was statistically reliable.

Response Modality

The average total times to complete the verbal and manual conditions were 35.85 (*SD* = 5.91) and 39.08 (*SD* = 5.58). The IVR took longer to navigate with the manual response modality, which was expected since it takes longer to press a key than to give a verbal command. The main effect of response modality $F(1, 37) = 17.81$, $p < .01$, MSE = 21.89, $\eta p^2 = .32$ was statistically reliable.

Navigation – Error Count

No main effects and/or interactions were found with the error count analysis.

Balance - Response Time

Mapping Type by Education by Replacement Type

An interaction between mapping type, education, and replacement type was obtained with the response time analysis for checking an account balance. As shown in Figure 15, performance across education level and replacement type for the direct mapping type was approximately the same. However, there were slight differences in performance across education level in the indirect condition as shown in Figure 16. The three-way interaction between mapping type, education, and replacement type was statistically significant, $F(1, 33) = 6.34$, $p = .02$, MSE = 6.67, $\eta p^2 = .16$. A simple post-hoc test shows an interaction between education and replacement type for the indirect condition $F(1, 33) = 4.26$, $p = .047$, MSE = 12.97, $\eta p^2 = .11$ but not for the direct condition $F(1, 33) = .176$, $p = .67$, MSE = 17.58, $\eta p^2 = .01$.

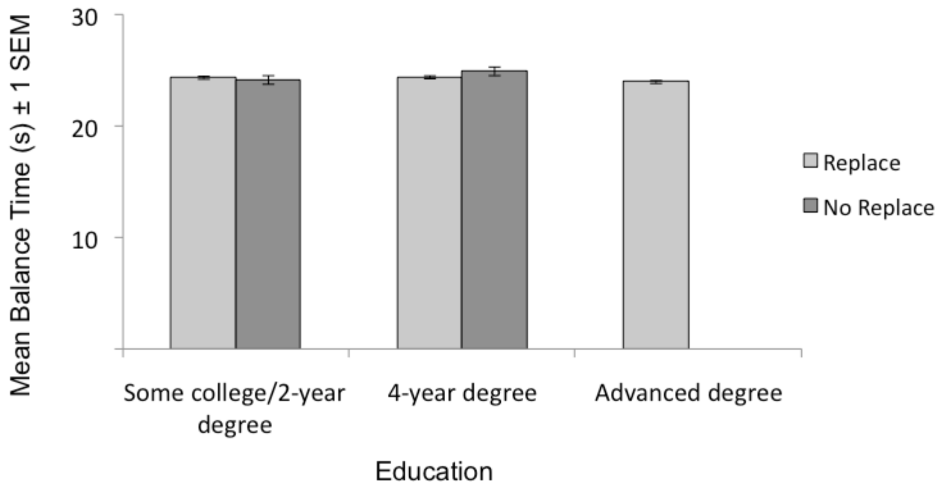


Figure 15. Interaction between education and replacement type for direct mapping type.

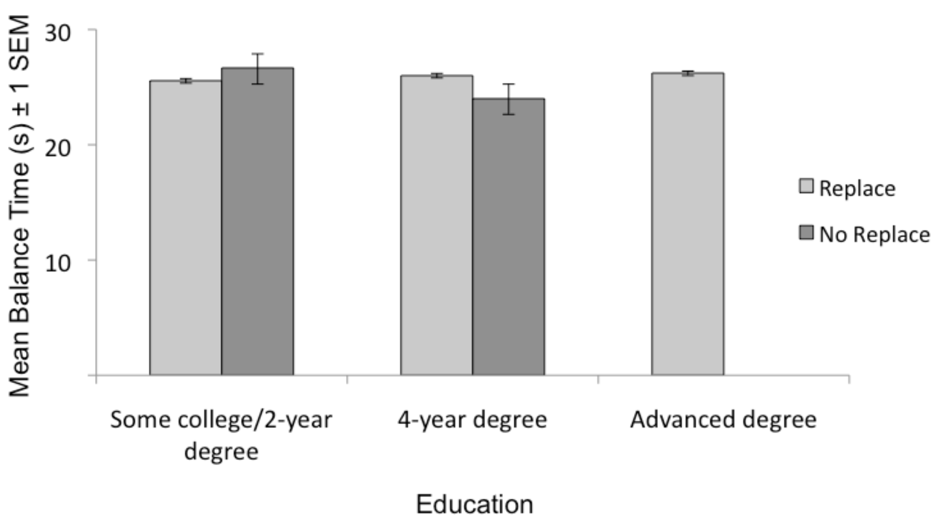


Figure 16. Interaction between education and replacement type for indirect mapping type.

Age

The average time taken to check an account balance for younger and older subjects was 24.38 (*SD* = 2.79) and 25.94 (*SD* = 3.08). Older subject were slower than younger subjects at checking their account balance. This result is an expected since prior research in aging shows that older adults are generally slower. The main effect of age $F(1, 33) = 8.31$, $p = .01$, MSE = 23.88, $\eta p^2 = .20$ was reliable.

Balance – Error Count

No main effects and/or interactions were found with the error count analysis.

Total – Response Time

Response Modality by Education by Replacement Type

An interaction between response modality, education, and replacement type was obtained with the response time analysis for the total call. Overall, as shown in Figure 17 and 18 subjects performed faster with the verbal response modality. Replacement type was approximately the same for each response modality and only slightly differed for subjects educated with a four-year degree as shown in Figure 18. The three-way interaction between response modality, education, and replacement type was statistically significant, $F(1, 33) = 5.86$, $p = .02$, MSE = 37.24, $\eta p^2 = .15$. A simple post-hoc test shows an interaction between response modality and replacement type for subjects with a four-year degree $F(1, 10) = 7.43$, $p = .02$, MSE = 39.40, $\eta p^2 = .43$ but not for subjects with some college/two-year degree $F(1,18) = .54$, $p = .47$, MSE = 32.75, $\eta p^2 = .03$ and for advanced degrees the result was inconclusive (the result was zero; in this case, six older subjects had advanced degrees, while only one younger subject had this type of degree).

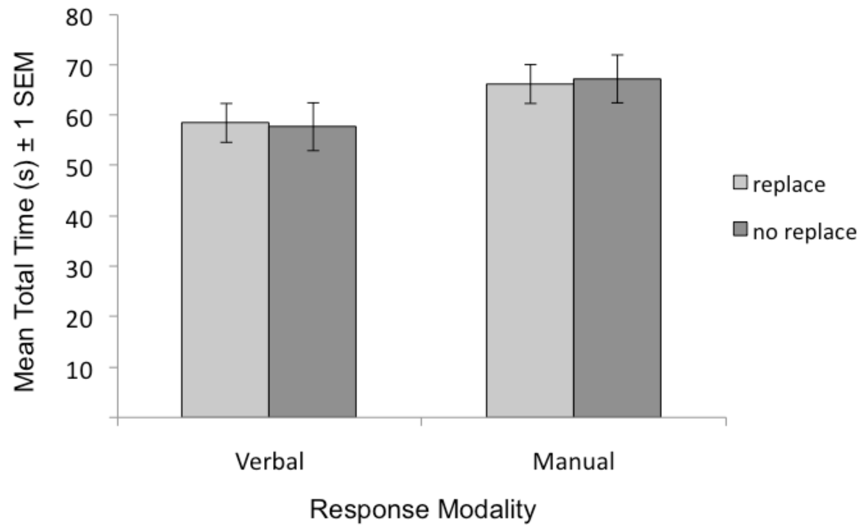


Figure 17. Interaction between response modality and replacement for some college/2-year degree.

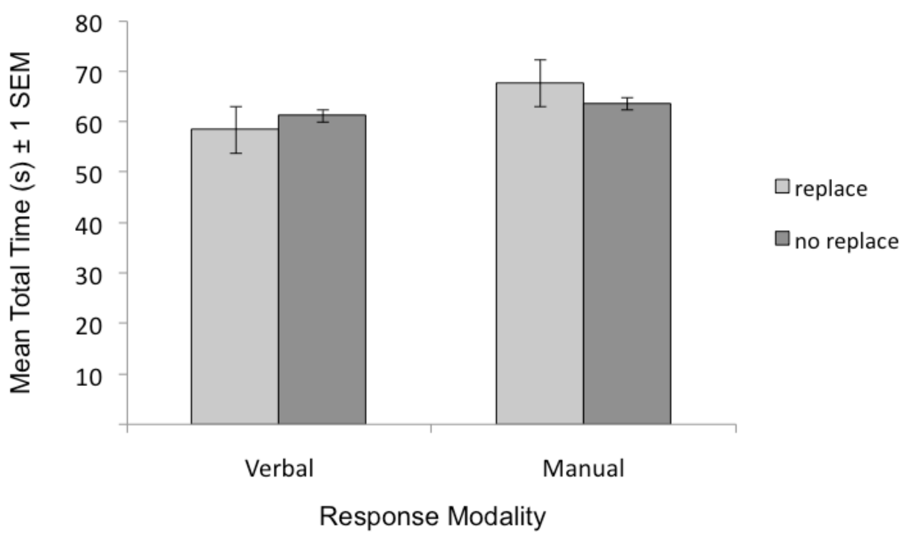


Figure 18. Interaction between response modality and replacement for 4-year degree.

Task

The mean total times to complete the banking and health tasks were 63.84 (*SD* = 6.77) and 61.11 (*SD* = 6.49). The banking task took slightly longer to complete. The main effect of task $F(1, 33) = 13.16$, $p < .01$, MSE = 26.16, $\eta p^2 = .29$ was reliable.

Response Modality

The mean total times to complete the verbal and manual conditions were 58.58 (*SD* = 6.79) and 66.37 (*SD* = 6.49). There is increased time for the manual response modality, which is expected since it takes longer to press a key than to give a verbal command. The main effect of response modality $F(1, 33) = 55.44$, $p < .01$, MSE = 37.24, $\eta p^2 = .62$ was reliable.

Mapping Type

The average total times to complete the indirect and direct conditions were 63.51 (*SD* = 6.84) and 61.44 (*SD* = 6.43). The indirect condition took slightly longer to complete than the direct condition. This result is expected since the indirect conditions involved inputting a value that was not memorized, which would take longer. The main effect of mapping type $F(1, 33) = 4.25$, $p = .047$, MSE = 40.85, $\eta p^2 = .11$ was reliable.

Total – Error Count

No main effects and/or interactions were found with the error count analysis.

DISCUSSION

As outlined previously, there were two hypotheses for how response modality (verbal or manual response) affects older and younger adults' performance when using an IVR. The first hypothesis was that the difference between high set (auditory stimulus - verbal response) and low set-level compatibility (auditory stimulus - manual response) is larger for older than younger adults in an IVR task (banking or healthcare). The second hypothesis was that the difference between mapping type (direct or indirect) and response modality would be larger for older than younger subjects in an IVR task. The results did not support these hypotheses.

Overall, subjects were faster navigating the IVR in the high set-level compatibility conditions. However, no effect of age was found when high set and low set compatibility were examined. This result is inconsistent with prior research that has shown that performance of older adults is more disrupted by incompatibility than younger adults (Doose et al., 2001; Emery et. al., 2006). Since speed of processing and working memory decline with age, it was expected that older adults would perform worse on low-set level compatibility tasks than younger adults. A potential explanation is the nature of the task used for this study was different from the tasks tested in previous research. The tasks used to test set-level compatibility in the literature were laboratory tasks that required very fast processing, subjects had to say a word or press a button very quickly across many trials, which was on the order of milliseconds. Since the speed of the task was so fast, small age differences like 50 milliseconds are important. For example in the Doose et al., (2001) study described earlier, the difference between older and younger adults on conceptual comparison task using the verbal response modality was 64 milliseconds. In the current study, fast processing was only required when subjects had to respond to a prompt by the IVR. Unlike the laboratory tasks, it was impossible to measure the latency from the start of stimulus (the verbal prompt with the menu options in this case) to when subjects used the IVR to make a response. As a result, deficits that were found in the literature could not be found using the IVR in this study and if there were any age differences they may only have constituted a small proportion of total time. Even if latency could be properly measured, real differences may exist but may not be large enough to cause a difference in actual performance. A reason is because users have a certain amount of control when using an IVR (barge-through menu options) that could possibly compensate for any deficits that may occur. The type of control refers to users having the choice to listen to all of the menu options or barging-through the menu when their desired option is heard. In all, the results indicate that there is no effect of set-level compatibility on age in an IVR task, which can possibly be attributed to task differences and other external factors. Performance can change across tasks and the main purpose of this research was to determine how it changes using the IVR.

A three-way interaction between age, response modality, and mapping type would have lent support for the second hypothesis. While older subjects were generally slower at checking their account balance and all subjects performed better in the direct mapping condition using the verbal response modality, no such three-way interaction was found. However, a two-way interaction between response modality and mapping type was found. Since it typically takes longer to input information manually than verbally, the two-way interaction was expected because it was assumed that the difference between the direct and indirect mapping type would be larger for the manual response modality. Furthermore, it was assumed that this difference would be larger for older than younger adults (response modality by mapping type by age), because it was believed that older adults would perform more slowly and error more frequently than younger adults in the indirect mapping type condition. The significant interaction between response modality and mapping type indicated that there was only a slight difference in mapping type for the verbal condition, and as expected the cost of mapping type was larger for the manual condition. This result makes sense because people tend to talk more slowly when verbally interacting information with an IVR, so their responses are more likely to be understood. As a result, in both mapping type conditions the information would have been input at the same pace using the verbal response modality. Also the indirect mapping type would have been faster with the verbal than manual response modality because the information inputted did not have to be mapped to the telephone keypad. Another likely reason why subjects were faster with the indirect mapping type using the verbal response modality was because when using the manual response modality, subjects had to refer back and forth between the index card (i.e., the index card containing the ten and four digit number) and the telephone keypad in order to input the numbers successfully. While the four digit number could be input by only referring to the index card once, the ten digit number had to be referred to several times. The reason for so many referrals was because there were no dashes between the ten digit number and most people have difficulty remembering more than six digits at a time. In all, the results indicate that while mapping type does have an effect on task performance, there is no difference in performance between older and younger adults.

One anomaly found in the results was the errors obtained when subjects navigated the IVR. It was expected that subjects would perform better (i.e., fewer errors and faster response times) in the high set-level compatibility conditions. However, as shown in Figure 9, the opposite appears to be true: subjects performed worse in the high set-level compatibility conditions when navigating the IVR. On further examination of these data it appears that most of the errors occurred in the healthcare condition and the types of errors obtained were invalid entry or no entry. In the high set-level compatibility condition, subjects were prompted to say a number that was associated with an menu option (e.g., for oncology, say six). Most of the invalid entry errors were obtained when subjects tried to speak the actual menu option (e.g., oncology) instead of speaking the number associated with the option (e.g., six). The no entry (silence) occurred when subjects did not respond to the IVR menu prompt. Subjects may have performed worse when performing the healthcare task because they were required to remember long, unfamiliar medical words to select a menu option (e.g., oncology and ophthalmology). Since the IVR menu used to navigate the system had no barge-through option, the long medical words may have made it harder for subjects to remember the correct menu option once they were prompted for a response. As a result of not remembering the menu option, subjects obtained a no entry error. Also, subjects may have been used to speaking the actual word (e.g., oncology) instead of a number associated with the menu option (e.g., six), which may have caused the invalid entry errors. In most IVRs users are allowed to say the actual word and use barge-through to avoid having to remember irrelevant menu options. The assumption is that subjects assumed made fewer errors in the banking task, because the words used for the menu options were shorter. However, while subjects made fewer errors in the banking task they were faster navigating the healthcare task, which may have been due to the subject's own mental model. Subjects were given a scenario that involved checking a savings account balance. On a few of the scenarios (see Appendix A), some subjects expressed that a checking account should be used instead and as a result they selected that particular incorrect option from the menu.

Failure to find evidence that set-level compatibility and mapping type cause differences in task performance between older and younger subjects in the present study may be accounted for by several potential factors. First, the aforementioned task differences between the literature and the current study (discussed previously) maybe one reason for the lack of age-related findings. Second, the IVR design may have been another determining factor, as it was specifically designed to address the cognitive needs of older adults, which may have helped compensate for any deficits that may have occurred. This is supported by the fact that seventy-eight percent of the older subjects rated the IVR systems in this study as easy to use. Although a small sample size was used for this study, the results showed no difference in IVR performance between older and younger adults. This suggests that the IVRs created for the study suppressed age effects. As a result of the study, four design guidelines were created that can be used to make IVRs that will better accommodate the needs of older adults.

- Use the best menu structure for your application.

  The type of menu structure selected for use in an IVR should be based on the type of application and capabilities of the target audience. There are two choices, broad or deep menu structure. As discussed previously, both have advantages and disadvantages. Weighing these choices is important because menu structure may help users navigate the IVR more successfully and lead to greater satisfaction. In the study, a broad menu structure was used because previous research has shown that broad-based menus are better for people with lower working memory capacity, which is common in older adults.

- Provide good error recovery.

  Even with the best possible IVR design, it is nearly inevitable some users will make errors. Therefore it is essential to incorporate good error recovery with your IVR design. Good error recovery involves three steps: (1) inform the user when an error is made and the type of error (e.g., "you entered a invalid account number"), (2) allow for recovery (e.g., "please re-enter your account number"), and (3) repeat until the error has been corrected. These error recovery steps were used in the IVR created for the study. However, given the time constraints of a laboratory study, subjects were only given three chances to successfully recover from an error.

- Use an excellent synthetic voice or voice actor.

  A common complaint about IVR systems is difficulty understanding the voice prompts. In terms of type of voice, there are two options: synthetic voice or hiring a voice actor. Synthetic voices have continued to improved over time and now sound more realistic than they did years ago. Although more expensive than a synthetic voice, a voice actor is nice to use because their voice can capture the correct tone (e.g., bubbly and cheerful versus serious and reserved) of a business. Whatever the voice type chosen, all prompts should be spoken in a clear and concise manner. A synthetic voice was used for the study. The QuickFuse platform used to create the IVR had over 50 voices from different genders and regions. Voices were sampled until the correct one that fit the profile of the businesses used in the IVR were found.

- Design prompts that are short and to the point.

  Another common complaint about IVR systems is that the prompts are too long. It is imperative that IVR prompts are designed to be concise and to the point. Longer prompts can make it difficult for subjects to navigate the system successfully, since subjects have to hold many items in memory. The prompts used in the study were short. In addition, each prompt in the IVR system was designed to have approximately the same number of words.

Third, context and experience may have played a role in why older adults did not perform differently from younger adults. In the present study, a basic research paradigm was put in an applied real-world context. As a result of this context, older adults were able to benefit from their previous experience with using an IVR and/or familiarity with a task domain (healthcare and banking) to perform better in the study. Previous research in the aging literature has shown that expertise may be able to improve performance on a task. Castel (2007) showed that older adults with an accounting background were able to recall numerical information better than older adults without the same background. In addition, older adults with the expertise actually outperformed younger adults without the same experience. The results from the Castel (2007) study might also explain why subjects performed better with one task domain over the other in the current study. It could be simply that participants had more expertise with one task domain over the other. In the study, the IVR systems for each task domain were designed exactly the same in terms of number of menu options, decision points and menu structure; the only difference was the task domain content. As a result, it was unexpected to find any effect of task. However, subjects were faster navigating the IVR and completing the total call in the healthcare condition. This result may suggest that IVR performance is influenced by prior real-world experience with a particular domain, which could indicate that task domain is at least partially independent of IVR design. This means that familiarity with a task domain could help compensate for a poor IVR design, or improve efficiency and effectiveness with a good IVR design.

As with any study, there were a few limitations that should be discussed:

IVR System

The QuickFuse application is a commercial Do-it-Yourself system that allows users to create their own IVR system. QuickFuse is mainly used for small-to-medium sized businesses and is not a research tool. As a result, some of the functionality required for a larger enterprise and for this type of research were missing. An activity report was one of the missing functionalities in the IVR system. An activity report maintains all of the actions performed by a user in a call to an IVR system, actions such as every key-press and menu option selected, along with associated time stamps. Activity reports also collects other helpful information, such as if and when information was retrieved from a database and how long someone took to select an option, which could have helped determine processing time. The lack of activity report left the study without information that would have been helpful for finer-grained analyses. Notes and video from the session were recorded to supplement some of the information that would have been provided by the activity report. The ability to create an IVR that solely accepted verbal or manual input was another missing functionality. The IVR accepted both types of input; observing the subjects and intervening if the subject used the wrong response modality handled this problem. Also, video from each session was reviewed to ensure that subjects used the correct response modality for the specified condition.

With any IVR system, speech recognition is not going to be 100% accurate; this was the case with the speech-recognition in the QuickFuse application. As a result, video and notes from each session were reviewed to ensure that errors were correctly attributed to the input made by the subject rather than the IVR system's failure to recognize a correct response.

Another issue was that in the QuickFuse database leading zeros were removed. As a result, subjects whose birth month began with a zero, as in the month of April, were instructed to input their birth month as “14” rather than “04.” This was an issue of concern because replacing the zero with a one made the direct mapping condition only half direct. Seventy-four percent of the subjects in the study made a replacement. Despite this issue, only a few people forgot to replace the zero with a one and there was approximately no difference in performance between replacement types. At various points in the experiment subjects were reminded to replace the zero with a one. Another issue was a response delay in the IVR system. The IVR response to a selection made by a user was slower for the verbal than the manual response modality. The response delays were subtracted from the response time variables. Researchers using these type of IVR systems should be mindful of system constraints such as these and how they can impact their dependent variables.

IVR Design

In an IVR, effective prompts help the user navigate the system and complete their desired task(s). The IVR prompts in the study were modeled after commercial banking and healthcare IVR prompts.

However, in most systems the user has the option to either verbally or manually enter their responses. An IVR was created for each response modality and task domain. In the manual response IVR, subjects were instructed to use the telephone keypad to press a number. In the verbal response IVR, subjects were instructed to say their response. The prompts for each response modality IVR were not equal; it takes more words for an IVR prompt to say “enter your response using your telephone keypad” versus “say your response now.” As a result, it is possible that the manual response modality took longer to complete simply because the prompts were longer. The prompts were also slightly unequal for the task domains. One way to handle this issue would have been to subtract the prompt times from the response time variables. However, as discussed in the methodology, half the prompts in the system were barge-through. Since the IVR system did not have an activity report log, it would have been almost impossible to get an accurate time of when the barge-through occurred. Using barge-through for only half of the prompts was a poor system design. However, barge-through was used because without it the system had trouble picking up responses that were entered immediately after a prompt. For this reason, most decision points had two parts: the menu, which was no barge-through and the prompt informing the subject to input their response as shown in Figure 23. The separate prompts possibly caused a usability issue. In the study, more than half of the subjects ended the call after their balance was heard despite being instructed to not terminate the call until they were instructed to by the IVR. It is unclear if this was due to a poorly designed prompt or reflects what happens in the real world, where people hang up after hearing their desired information. After the prompt was heard, subjects would press or speak their response before being instructed to and as a result their responses were sometimes not picked up by the system. The prompts could have been better designed to prevent this problem.

Future research should take the limitations of this study into consideration, as well as investigate how different task domains and menu structure affect task performance for older and younger adults when using an IVR. Given the fact that there was a difference in performance based on task domain, it would be interesting to examine how different domains affect task performance with an IVR. This research would allow the field to determine if and how the effects of response modality and mapping type vary by task. Also this would enable the evaluation of whether familiarity with a task domain could help compensate for a poor IVR design or improve efficiency and effectiveness for a good IVR design. Next, menu structure in an IVR should be examined. It would be interesting to research how differences in menu structure affect task performance, response modality and mapping type.

The purpose of this exploratory research was to examine the effect of response modality and mapping type on task performance for older and younger adults when using an IVR. Most of the research conducted on the effect of response modality (specifically, set-level compatibility) on aging has involved contrived laboratory tasks. Therefore, it was unclear if the results from those studies could be generalized to a task that was more representative of the real world; the results of this study provide some insight. The results suggest that set-level compatibility does not affect task performance of older adults differently than younger adults with an IVR. In addition, the research shows that while people are faster inputting a value they have memorized (direct mapping type), the response modality used can affect performance. Overall, the primary contributions of this research are the design guidelines that were created to develop an IVR that suppressed age effects. These guidelines have practical applications; they should not just be of interest to IVR designers, but also to the larger human factors community as the role of technology continues to evolve and the population continues to age.

REFERENCES

Baddeley, A. D. (1999). *Essentials of human memory*. Hove, England: Psychology Press.

Borgman, C. (1986). The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Man-Machine Studies*, *24*(1), 47-64. doi: 10.1016/S0020-7373(86)80039-6

Brandt, J. (2008). Interactive voice response interfaces. In P. Kortum (Author), *HCI beyond the GUI: Design for haptic, speech, olfactory and other nontraditional interfaces* (pp. 229-265). Amsterdam: Elsevier/Morgan Kaufmann.

Castel, A. D. (2007). Aging and memory for numerical information: The role of specificity and expertise in associative memory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *62*(3), P194-P196.

Doose, G., & Feyereisen, P. (2004). Stimulus–response translation in manual and vocal modality: Is age‑related slowing influenced by changes in working memory functioning? *European Journal of Cognitive Psychology*, *16*(4), 597-624. doi: 10.1080/09541440340000394

Dulude, L. (2001). *Usability of interactive voice response systems in real-life tasks performed by old and young women* (Unpublished master's thesis). Carleton University.

Ekstrom, R. B. (1996). *Manual for kit of factor referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Emery, L., Myerson, J., & Hale, S. (2006). Saying versus touching: Age differences in short-term memory are affected. *Journal of Gerontology*, *61*(6), 366-368.

Engelbeck, G., & Roberts, T. (1990). *The effects of several voice-menu characteristics on* *menu-selection performance* (pp. 50-63, Rep. No. ST0401). US West Advanced Technologies.

Evans, R. (2009). *The impact of voice Characteristics on User Response in an interactive voice response system* (Unpublished master's thesis). Rice University, Houston, TX.

Fairbrother, J. T. (2010). *Fundamentals of motor behavior*. Champaign, IL: Human Kinetics.

Gardner-Bonneau, D. (1999). Guidelines for speech-enabled IVR application design. In *Human factors and voice interactive systems* (pp. 146-162). Boston: Kluwer Academic.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11-26. doi: 10.1080/17470215208416600

Hiremath, V., Fanjoy, R., Feyen, R., & Young, J. (2009). Comparison of pilot recovery and response times in two types of cockpits. In M. J. Smith (Author), *Human interface and the management of information: Symposium on Human Interface 2009, held as part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009 ; proceedings.* Berlin: Springer.

Hommel, B., & Prinz, W. (1997). *Theoretical issues in stimulus–response compatibility*. Amsterdam: Elservier.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188-196. doi: 10.1037/h0056940

Kantowitz, B. H., Triggs, T. J., & Barnes, V. E. (1990). Stimulus-response compatibility and human factors. In R. W. Proctor (Author), *Stimulus-response compatibility:*

An integrated perspective (pp. 365-388). Amsterdam: North-Holland.

Kline, D. W., & Scialfa, C. T. (1997). Sensory and perceptual functioning: Basic research and human factors implications. In A. D. Fisk & W. A. Rogers (Authors), *Handbook of human factors and the older adult* (pp. 327-338). San Diego: Academic Press.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility--A model and taxonomy. *Psychological Review*, *97*(2), 253-270. doi: 10.1037//0033-295X.97.2.253

Kornblum, S. (1990). S-S and S-R compatibility effects as processing consequences of overlap on relevant and irrelevant dimensions. *Bulletin of the Psychonomic Society*, 28, 502.

Lohman, D. F. (1996). Spatial ability and G. In I. Dennis & P. Tapsfield (Authors), *Human abilities: Their nature and measurement* (pp. 97-116). Mahwah, NJ: Lawrence Erlbaum Associates.

Meiran, N., & Gotler, A. (2001). Modelling cognitive control in task switching and ageing. *European Journal of Cognitive Psychology*, *13*(1-2), 165-186. doi: 10.1080/09541440126280

Milham, M. P., Erickson, K. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., & Cohen, N. J. (2002). Attentional Control in the Aging Brain: Insights from an fMRI Study of the Stroop Task. *Brain and Cognition*, *49*(3), 277-296. doi: 10.1006/brcg.2001.1501

Pak, R., Rogers, W. A., & Fisk, A. D. (2006). Spatial Ability Subfactors and Their Influences on a Computer-Based Information Search Task. *Human Factors: The*

Journal of the Human Factors and Ergonomics Society, *48*(1), 154-165. doi: 10.1518/001872006776412180

Peterson, K. D. (2000). *Business telecom systems: A guide to choosing the best technologies and services*. New York, NY: CMP Books.

Proctor, R. W., & Vu, K. -P. L. (2001). Stimulus-response compatibility in interface design. In M. J. Smith (Author), *Usability evaluation and interface design: Cognitive engineering, intelligent agents and virtual reality : Volume 1 of the proceedings of HCI International 2001...* (pp. 1368-1372). Mahwah: Lawrence Erlbaum.

Proctor, R. W., & Vu, K. -P. L. (2008). Human information processing: An overview for human-computer interaction. In A. Sears & J. A. Jacko (Authors), *Human-computer interaction.* Boca Raton, FL: CRC Press.

Proctor, R. W., & Wang, H. (1997). Set- and Element-Level Stimulus-Response Compatibility Effects for Different Manual Response Sets. *Journal of Motor Behavior*, *29*(4), 351-365. doi: 10.1080/00222899709600021

Proctor, R. W., Lu, C., Van Zandt, T., & Weeks, D. J. (1994). Affordances, codes, and decision processes: A response to Michaels (1993). *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 452-455. doi: 10.1037//0096-1523.20.2.452

Proctor, R. W., Vu, K. L., & Pick, D. F. (2005). Aging and Response Selection in Spatial Choice Tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(2), 250-270. doi: 10.1518/0018720054679425

Profile america facts for features, older american month. (2011, March 23). *Facts for*

Features: Older Americans Month: May 2011. Retrieved from http://www.census.gov/newsroom/releases/archives/facts_for_features_special_editions/cb11-ff08.html

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763-776. doi: 10.1037//0012-1649.27.5.763

Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sharit, J., Hernández, M. A., Czaja, S. J., & Pirolli, P. (2008). Investigating the Roles of Knowledge and Cognitive Abilities in Older Adult Information Seeking on the Web. *ACM Transactions on Computer-Human Interaction*, *15*(1), 1-25. doi: 10.1145/1352782.1352785

Simon, J. R. (1967). Choice Reaction Time as a Function of Auditory S-R Correspondence, Age and Sex. *Ergonomics*, *10*(6), 659-664. doi: 10.1080/00140136708930919

Smither, J. A. (1993). Short term memory demands in processing synthetic speech by old and young adults. *Behaviour & Information Technology*, *12*(6), 330-335. doi: 10.1080/01449299308924397

Verhaeghen, P., Marcoen, A., & Gossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journals of Gerontology*, *48*(4), 157-171.

Wang, H., & Proctor, R. W. (1996). Stimulus-response compatibility as a function of stimulus code and response modality. *Journal of Experimental Psychology:*

Human Perception and Performance, *22*(5), 1201-1217. doi: 10.1037//0096-1523.22.5.1201

APPENDIX A

Block 1 (B1): (Auditory-Manual) – Direct Condition

Healthcare task (8 steps) - Patient Billing
B1: Scenario 1 – You recently had a surgical procedure done at Big Hospital and would like to find out the amount you owe. Call Big Hospital at 1-855-346-2291 to reach the Billing department to determine the amount of your bill. Navigate the system and then **enter** the following into the telephone keypad when instructed:

- Your medical number is your 10 digit phone number
- Your group number is the month and day of your birthday. **Note: if your birth month has a 0 (zero) for the first number, replace it with a 1 (one).**

Please exit the system when instructed to and then record the amount you owe on the answer sheet.

Banking task (8 Steps) – Checking account
B1: Scenario 2 – You are thinking about buying a new TV and would like to make sure your checking account has enough money. Call Big Bank at 1-855-351-4897 to check your balance. Navigate the system and then enter the following into the telephone keypad when instructed:

- Your medical number is your 10 digit phone number
- Your group number is the month and day of your birthday. **Note: if your birth month has a 0 (zero) for the first number, replace it with a 1 (one).**

Please exit the system when instructed to and then record the amount you owe on the answer sheet.

Block 2 (B2): (Auditory-Manual) – Indirect Condition

Healthcare task (8 steps) - Patient Billing
B2: Scenario 1 – You have started a new job and need to change to a new medical provider. However, first you must pay the balance at your old medical provider. Call Big Hospital at 1-855-346-2291 to reach the Billing department to determine the amount of you owe. Navigate the system and then enter the following into the telephone keypad when instructed:

- Your medical number is: **2576912983**

- Your group number is: **1359**

Please exit the system when instructed to and then record the amount you owe on the answer sheet.

Banking task (8 Steps) – Checking account
B2: Scenario 2 – You would like to purchase a designer dress for your niece for her prom, but would like to determine how much is in your checking account first. Call Big Bank at 1-855-351-4897 to check your balance. Navigate the system and then **enter** the following into the telephone keypad when instructed:

- Your account number is: **7594262354**

- Your ID number is: **2351**

Please exit the system when instructed to and then record the amount on you owe on the answer sheet.

Block 3 (B3): (Auditory-Verbal) – Direct Condition

Healthcare task (8 steps) - Patient Billing
B3: Scenario 1 – You recently received an enormous bill from your health insurance provider, which you feel is an error. Now, you would like to determine if the balance on the bill matches the amount owed at the hospital where the services were performed. Call Big Hospital at 1-855-351-4900 to reach Billing to determine the amount of your bill. Navigate the system and **say** the following information when instructed:

- Your medical number is your 10 digit phone number

- Your group number is the month and day of your birthday. **Note: if your birth month has a 0 (zero) for the first number, replace it with a 1 (one).**

Please exit the system when instructed to and then record the amount you owe on the answer sheet.

Banking task (8 Steps) – Checking account
B3: Scenario 2 – You received your electricity bill and the summer heat (due to constantly running AC) has caused it to skyrocket. Your electricity bill is automatically taken from your checking account at the beginning of every month. Check your checking account balance to determine if you have money in your

account. Call Big Bank at 1-855-351-4906 to check your balance. Navigate the system and **say** the following information when instructed:

- Your account number is your 10 digit phone number
- Your ID number is the month and day of your birthday. **Note: if your birth month has a 0 (zero) for the first number, replace it with a 1 (one).**

Please exit the system when instructed to and then record the amount you owe on the answer sheet.

Block 4 (B4): (Auditory-Verbal) – Indirect Condition

Healthcare task (8 steps) - Patient Billing

B4: Scenario 1 – You are responsible for paying half of the medical bill for a minor accident that occurred on your property. You now have to find out the amount of your bill, so you can tell the other responsible party how much they owe. Call Big Hospital at 1-855-351-4900 to reach Billing to determine the amount of your bill. Navigate the system and **say** the following information when instructed:

- Your medical number is: **9342712825**
- Your group number is: **2567**

Please exit the system when instructed to and then record the amount you owe on the answer sheet.

Banking task (8 Steps) – Checking account
B4: Scenario 2 – Your have been using your checking account to save for a trip around the world. You think that you are finally ready to purchase the tickets, which is $5000 dollars. However, you must first check your account balance to determine if you have enough money. Call Big Bank at 1-855-351-4906 to check your balance. Navigate the system and say the following information when instructed:

- Your account number is: 3595278932
- Your ID number is: 8232

Please exit the system when instructed to and then record the amount on the answer sheet

APPENDIX B

Subject No.______

IVR Task Answer Sheet

Please write down answers on the available space.

B1:

Health care task
How much do you owe at Big Hospital? ________________

Banking task
What is your checking account balance at Big Bank? ________________

B2:

Banking task
What is your checking account balance at Big Bank? ________________

Health care task
How much do you owe at Big Hospital? ________________

B3:

Health care task
How much do you owe at Big Hospital? ________________

Banking task
What is your checking account balance at Big Bank? ________________

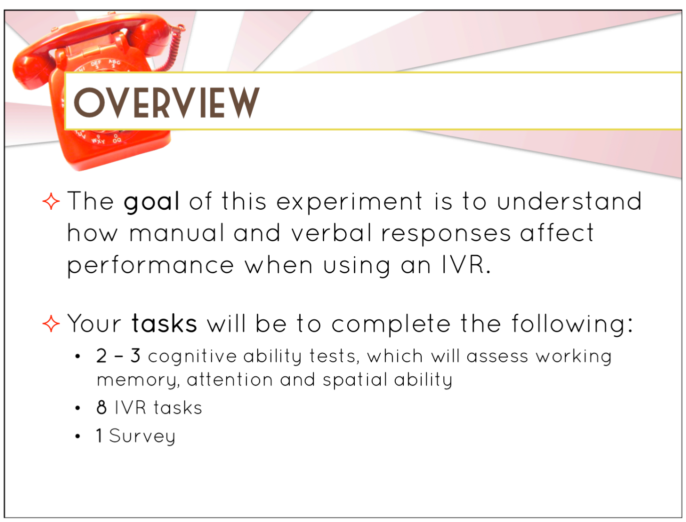
B4:

Banking task
What is your checking account balance at Big Bank? ________________

Health care task
How much do you owe at Big Hospital? ________________

APPENDIX C

Instruction slides read to participants during the experiment



LET'S BEGIN!

The cognitive ability tests will be administered now.

IVR TASKS

- You will be presented with a series of 8 scenarios that involve you:
  - Calling a number
  - Navigating a Interactive Voice Response system with a verbal or manual response
  - Obtaining information and writing it on an answer sheet, which is located on your clipboard.
    - *About the task sheet, you will match the scenario given to you with a location on your answer sheet. For example, if I gave you B4:Scenario 1, you would put your answer in this spot*

IVR TASKS

- In four of the tasks, you will use your personal:
  - 10 digit phone number you provided to the experimenter.
    - *For Example*, 2062813231
  - 4 digit month and day of your birth
    - *For example*, July 4th = 0704
  - According to my records, these numbers are:
- In four of the tasks, you will use numbers given to you

IVR RULES

- Birthdays
  - If your birth month begins with a **0**, you must replace it with a **1**.
    - For example, someone born on April 4th would enter *1404* instead of *0404*
- Please pronounce all zeros, as zeros.
- No barge-through option
  - You must listen to all the options before making a selection
  - After the options have been presented, the IVR will tell you to when to make your selection

IVR RULES

- Please do not hang-up the call until the system tells you it is hanging up
- You must dial **9** before making a phone call
- Warning! All the problems with the system have not been worked out yet. You may encounter a busy signal.

Do you have any questions? Please ask before we began the tasks.

LET'S BEGIN!

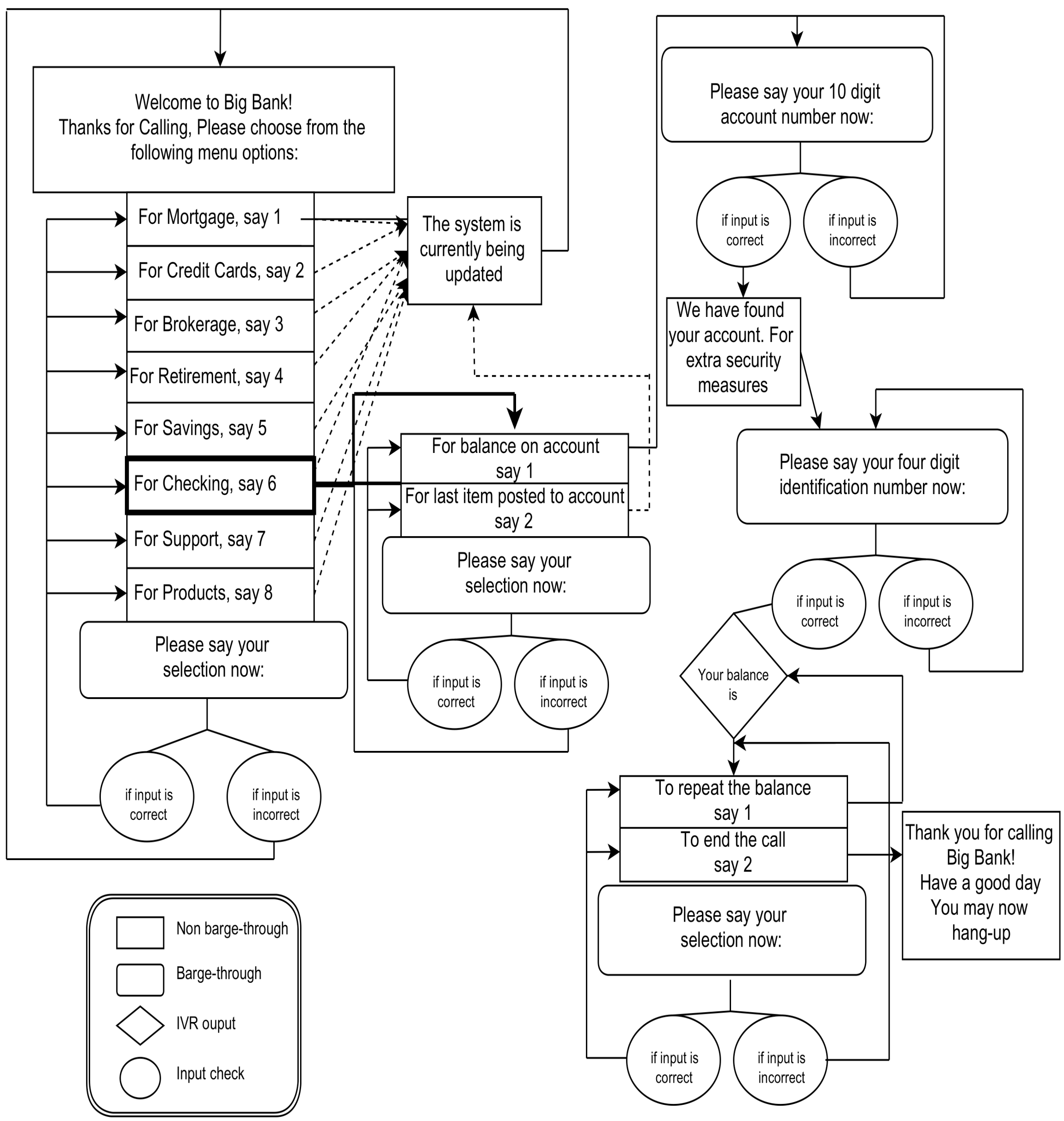
The IVR tasks will begin now.

THANK YOU – DEBRIEF

Thank you for participation, patience and time. The data we obtained from you will help us understand how performance of older and younger people differ on tasks that use manual and verbal responses.

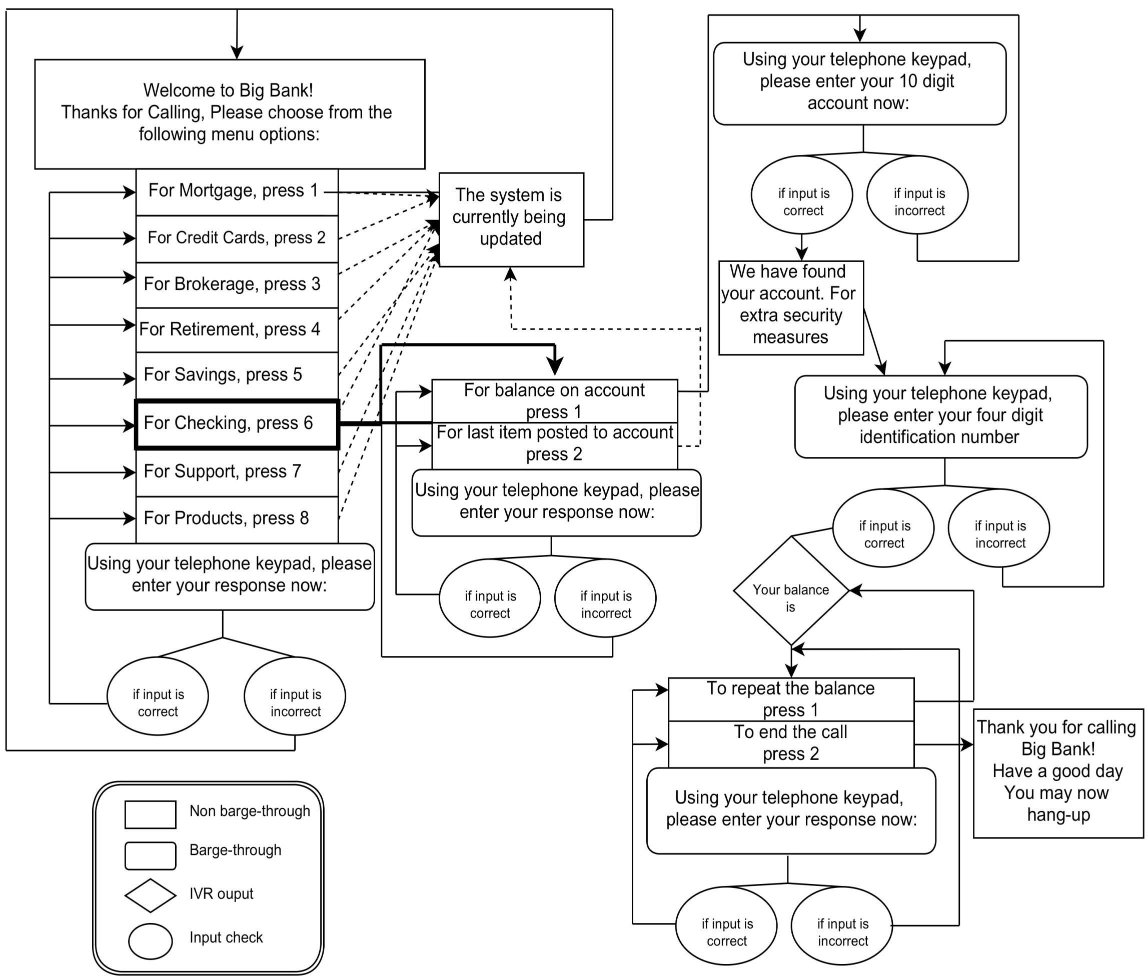
APPENDIX D

Bank-Verbal: Interactive Voice Response (IVR) System



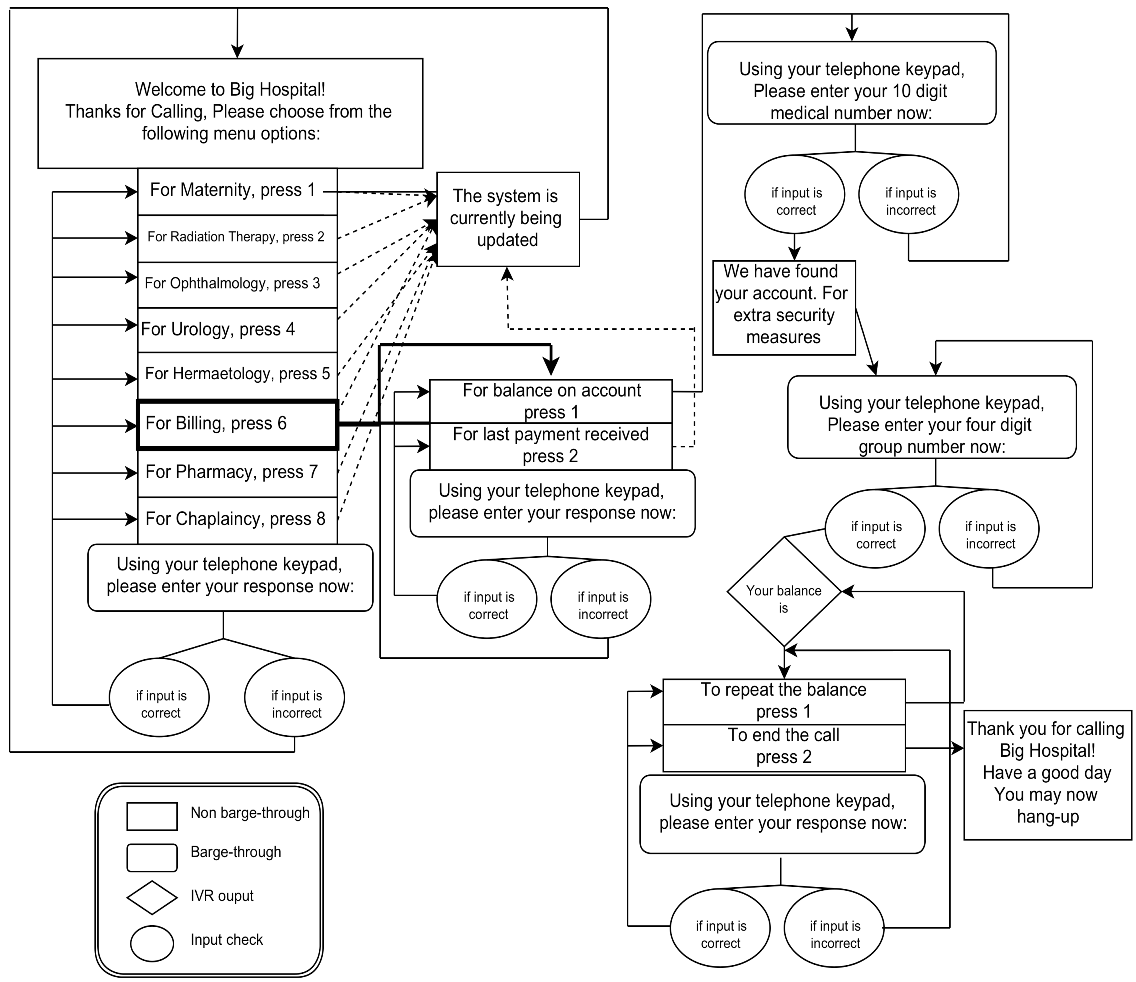
APPENDIX E

Bank-Manual: Interactive Voice Response (IVR) System



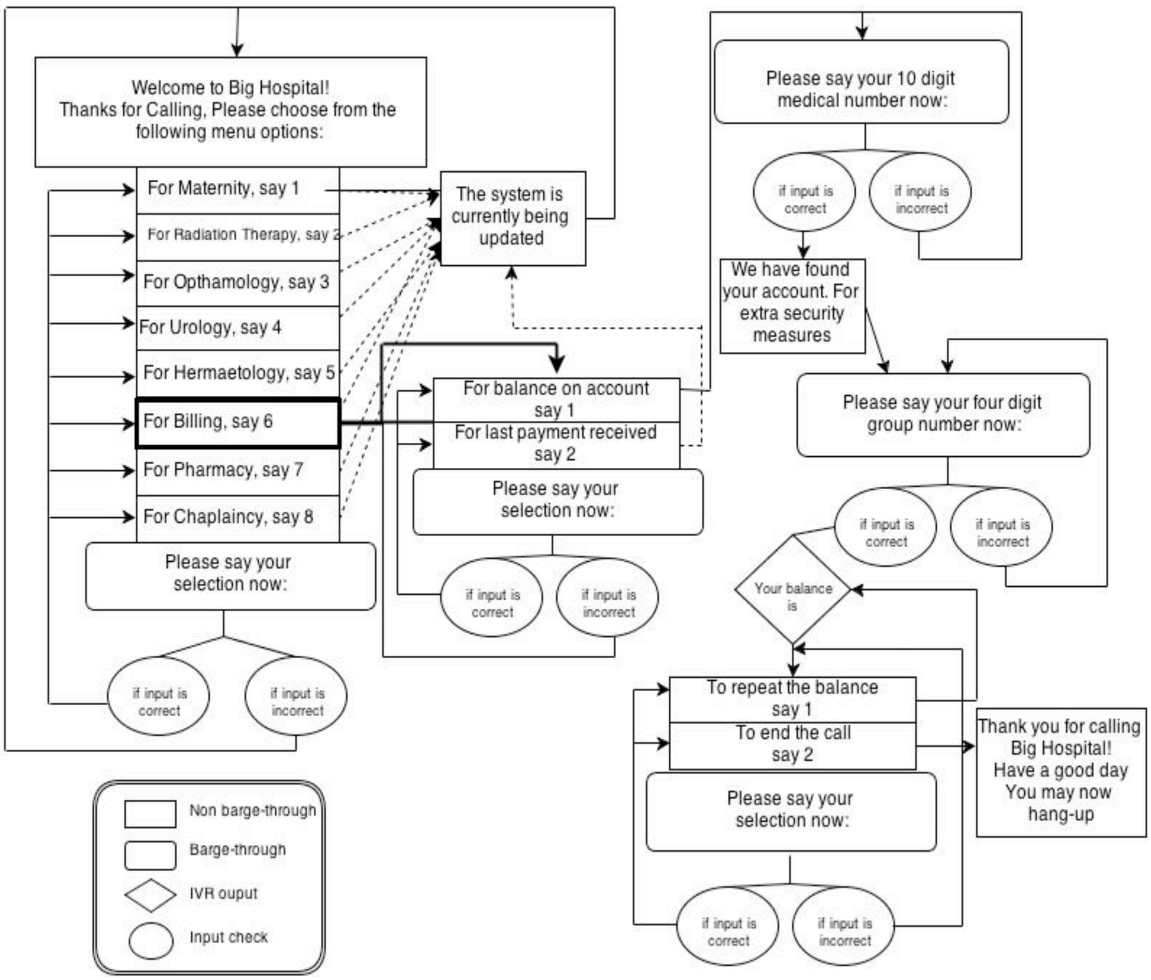
APPENDIX F

Healthcare-Manual: Interactive Voice Response (IVR) System



APPENDIX G

Healthcare-Verbal: Interactive Voice Response (IVR) System



APPENDIX H
Error Rates

Table H1

Mean Top and Second-Level Menu Error Rates

Top-Level	Second-Level
0.46%	1.86%

Table H2

Mean Top and Second-Level IVR Menu Error Rates for each Condition

Task Domain/ Response Modality/ Mapping Type	Top-Level	Second-Level
Banking-Verbal-Direct	0.33%	2.32%
Banking-Verbal-Indirect	0.65%	0
Banking-Manual-Direct	0.33%	1.16%
Banking-Manual-Indirect	0	2.33%
Healthcare-Verbal-Direct	1%	2.33%
Healthcare-Verbal-Indirect	0.67%	5.81%
Healthcare-Manual-Direct	0	1.16%
Healthcare-Manual-Indirect	0.33%	0

The error rates were calculated using the equation: sum of errors / opportunity for errors.In the top-level menu there were seven opportunities for errors. The second-level menu had two opportunities for errors.

APPENDIX I

Prompt Times and System Delays

Table I1

Total time for prompts in each condition

Task Domain/ Response Modality/	Prompt Times (*s*)
Banking-Verbal	43
Banking-Manual	47
Healthcare-Verbal	40
Healthcare-Manual	49

Due to task domain content and response modality wording, the total time for prompts are different in each condition. It is worth noting that the time reported is approximate, the time was obtained from reviewing videos from ten people in the study. The video does not report milliseconds.

Table I2

Response delays in the IVR system subtracted from response time variables

Task Domain/ Response Modality	Balance Time (*s*)	Navigation Time (*s*)	Total Time (*s*)
Banking-Verbal	10	12	22
Banking-Manual	2	3	5
Healthcare-Verbal	10	12	22
Healthcare-Manual	2	3	5

The response time variables were adjusted to accommodate a response delay in the IVR system. Response delays were subtracted from the response time variables.

APPENDIX J

Significant four and five way interactions

Table J1

Significant four and five-way interaction for the response time analysis for checking balance

Interaction (s)	Statistics
Response Modality by Mapping Type by Education by Replacement Type	$F(1, 33) = 4.91$, $p = .034$, MSE = 8.37
Tasks by Response Modality by Mapping Type by Education	$F(1, 33) = 3.49$, $p = .042$, MSE = 3.49

Table J2

Significant four-way interactiosn for the response time analysis for the total call

Interaction (s)	Statistics
Tasks by Education by Age by Replacement Type	$F(1, 33) = 5.77$, $p = .022$, MSE = 26.16
Tasks by Mapping Type by Age by Replacement Type	$F(1, 33) = 4.19$, $p = .049$, MSE = 31.58

APPENDIX K

#_____

Background and General IVR Survey

1. Age: _____

2. Gender: ___ Male ___ Female

3. Highest level of education?
   - ___ Less than high school
   - ___ High School GED
   - ___ Some College
   - ___ 2-Year College Degree
   - ___ 4-Year College Degree
   - ___ Master's Degree
   - ___ Doctoral Degree
   - ___ Professional Degree (MD, JD)

4. Please specify your race?
   - ___ American Indian or Alaskan Native
   - ___ Asian
   - ___ Black or African American
   - ___ Native Hawaiian or Other Pacific Islander
   - ___ White

5. Do you have normal or corrected to normal vision? ____ No ____ Yes

6. Are you left or right handed? ___ Right ___ Left ___ Ambidextrous

7. Are you a native English speaker? ___ No ___ Yes
   a. If no, what is your native language? _______________

8. Can you touch type? ____ No ____ Yes

9. How many hours per week do you use a computer?
   - ___ less than 5 hours
   - ___ between 5 and 10 hours
   - ___ between 10 and 20 hours
   - ___ between 20 and 30 hours
   - ___ between 30 and 40 hours
   - ___ over 40 hours

10. Of those hours, how many are spent on the internet?
    - ___ less than 5 hours
    - ___ between 5 and 10 hours
    - ___ between 10 and 20 hours
    - ___ between 20 and 30 hours

____ between 30 and 40 hours
____ over 40 hours

11. How many playing computer games?
    ____ less than 5 hours
    ____ between 5 and 10 hours
    ____ between 10 and 20 hours
    ____ between 20 and 30 hours
    ____ between 30 and 40 hours

12. Prior to your current level of computer usage, how many hours per week did you use a computer?
    ____ less than 5 hours
    ____ between 5 and 10 hours
    ____ between 10 and 20 hours
    ____ between 20 and 30 hours
    ____ between 30 and 40 hours

13. For how many years have you been at your current level of computer usage?
    ____

14. Please rate your level of computer expertise (1 = novice, 10 = expert)
    1 2 3 4 5 6 7 8 9 10

15. Which operating system do you use most frequently?
    ____ Windows
    ____ Macintosh
    ____ Linux/Unix
    ____ Other: ________________________

16. Which operating system do you prefer?
    ____ Windows
    ____ Macintosh
    ____ Linux/Unix
    ____ Other: ________________________

17. If you have a preferred search engine, which is it?
    ____

18. Do you have a webpage? _____ No _____ Yes

19. Which of these activities do you use a computer for? Check all that apply.

____ Word Processing (e.g. Microsoft Word)
____ Programming (e.g. Java, C++, Scheme)
____ Web design
____ Personal Finance (e.g. Quicken, Turbo Tax)
____ Games
____ Music
____ Multimedia (e.g encyclopedias; interactive CDs)
____ Spreadsheet management (e.g. Microsoft Excel)
____ Data Analysis (e.g. SAS, SPSS)

20. Which of the following activities do you use the World Wide Web for? Check all that apply.

____ Downloading Music
____ Shopping
____ Auctions
____ Finding Maps
____ Web Games
____ Webcast Videos
____ Weather forecast
____ News
____ Reference
______ Sports Scores
______ Internet Phone
______ Instant Messaging
______ Chat Rooms
______ E-mail
______ Blogging
______ Stock Trading
______ Dating

Prior Experiences with Interactive Voice Response System (IVR)

21. Have you ever used an IVR system in the past? ___ No ___ Yes

22. How many times in the past month have you used an IVR?

____ 1
____ 2
____ 3
____ 4
____ Other: ______________________

23. In the past, what type of the business have you used an IVR for? Check all that apply.

____ Banking
____ Utilities
____ Health care
____ Retail & Entertainment
____ Airline Travel
____ Weather forecast
____ Other: ______________________

Experience with experiment

24. How would you rate the difficulty of this experiment?
    a. ____ Very easy
    b. ____ Easy
    c. ____ Moderate
    d. ____ Difficult
    e. ____ Very Difficult

25. Use the following scale to rate the remaining questions:

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	agree	Strongly Agree

26. The IVR systems used in this study were easy to use? Check one that applies.
    ____ 1
    ____ 2
    ____ 3
    ____ 4
    ____ 5

27. The voice used to deliver the auditory prompt in the IVR systems was clear? Check one that applies.

    ____ 1
    ____ 2
    ____ 3
    ____ 4
    ____ 5

28. I understood the menu structure of the IVR system?
    Check one that applies.

    ____ 1
    ____ 2
    ____ 3
    ____ 4
    ____ 5

29. I thought some of the menu choices were frustrating?
    ____ 1
    ____ 2
    ____ 3
    ____ 4
    ____ 5

30. I had to concentrate really hard to complete each task?
    ____ 1
    ____ 2
    ____ 3
    ____ 4
    ____ 5