

**Interactive Voice Response Systems and Older Adults: Examination of the Cognitive
Factors Related to Successful IVR Interaction and Proof-of-Principle of IVR
Administration and Scoring of Neuropsychological Tests**

Delyana Ivanova Miller

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the
PhD degree in Clinical Psychology

Department of Psychology
Faculty of Social Sciences
University of Ottawa

© Delyana Ivanova Miller, Ottawa, Canada 2013

I dedicate this journey to my Mother

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Dr. Claude Messier for his exceptional mentorship, unwavering support and outstanding dedication to my learning experience and to this project. His unparalleled passion for knowledge, energy, understanding, patience and guidance made this experience one of the most inspiring intellectual pursuits of my graduate studies. I owe my aspiration for deeper understanding of complex research endeavours to his gentle resistance and creative outlook, which invited me to question many of my personal and intellectual convictions. I will always be grateful for his personal kindness and unswerving guidance.

To Dr. Claude Lamontagne, I would like to express my gratitude for his exceptional pedagogical dedication, support and encouragement during my entire university experience. His enthusiasm for knowledge and enquiring attitude taught me that asking difficult questions is more important than finding the answers. To him I owe many inspiring intellectual experiences and my aspiration for teaching.

I would also like to extend my gratitude to Vincent Talbot and the team at TelAsk Technologies for their strong commitment to this project and for generously offering their invaluable expertise. Without their support this undertaking would not have been possible.

I would also like to express my appreciation to the rest of my thesis committee members, Patrick Davidson, Kate Bielajew and Sylvain Gagnon for their interest and valuable input in this project and for making me feel welcome to turn to them for guidance whenever needed.

I would like to acknowledge my research assistants, Halina Bruce, France Aube, Hillary Kitchener, Ashley Charbonneau and Nigel McKechnie for their dedication and hard

work during the various stages of this project. Their commitment and enthusiasm were contagious and they undoubtedly made this journey an exceptional experience.

There are no words to fully express my gratitude to my mother, who had given me everything that matters the most. I am forever grateful to her unconditional love, positive attitude, tenacious personality and unwavering believe in my abilities. Because of her, I possess the self-confidence and perseverance that guided me throughout my journey in graduate school.

Finally, I wish to express my special thanks to my husband, Gary Miller, for his strong presence by my side in the last nine years. His love, support, and dependability have made all the difference for me being able to reach the end of this at times tumultuous journey. I would always be grateful for his understanding and genuine caring.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	III
TABLE OF CONTENTS	V
GENERAL ABSTRACT	1
GENERAL INTRODUCTION.....	3
<i>DEFINITION OF IVR SYSTEMS</i>	3
<i>IVR APPLICATIONS</i>	3
<i>IVR SYSTEMS AND OLDER PEOPLE</i>	6
<i>THE PRESENT STUDY</i>	11
<i>ORIGINALITY OF THE PROJECT</i>	13
IMPROVING OLDER ADULTS' EXPERIENCE WITH INTERACTIVE VOICE RESPONSE SYSTEMS	15
ABSTRACT.....	16
INTRODUCTION	17
MATERIALS AND METHODS	18
<i>Procedure</i>	19
RESULTS.....	20
<i>Demographics</i>	20
<i>First Focus Group Analyses</i>	20
<i>Second Focus Group Analyses</i>	23
DISCUSSION	24
REFERENCES	27
PREDICTORS OF SUCCESSFUL COMMUNICATION WITH INTERACTIVE VOICE RESPONSE SYSTEMS IN OLDER PEOPLE	30
ABSTRACT.....	31
INTRODUCTION	32
METHOD.....	35
<i>Participants</i>	35
<i>Apparatus</i>	37
<i>Procedure</i>	37
<i>Statistical Analyses</i>	39
RESULTS.....	40
<i>Measures That Predict the Number of IVR Tasks Completed</i>	40
<i>Measures That Predict the Level Achieved Within the IVR Tasks</i>	41
<i>Measures That Predict the Number of Errors Made During the IVR Tasks</i>	42
DISCUSSION	44
REFERENCES	49
TABLE 1.	52
TABLE 2.	53
TABLE 3.	54
TABLE 4.	55
OLDER PEOPLE'S ATTITUDES TOWARD INTERACTIVE VOICE RESPONSE SYSTEMS	57
ABSTRACT.....	58
INTRODUCTION	59
METHOD.....	63
<i>Participants</i>	63
<i>Measures</i>	64
<i>Apparatus</i>	65
<i>Procedure</i>	65
<i>Analyses</i>	67

RESULTS.....	68
<i>Success in Completion of the Four IVR Tasks</i>	68
<i>Participants' Perceptions of the IVR Systems</i>	69
<i>System Usability Analyses</i>	70
DISCUSSION	71
REFERENCES	74
TABLE 1.	76
APPENDIX A.....	77
APPENDIX B.....	78
ADMINISTRATION OF A VERBAL FLUENCY TASK AND THE DIGIT SPAN TASKS USING INTERACTIVE VOICE RESPONSE TECHNOLOGY: VALIDATION AND LIMITATIONS.....	85
ABSTRACT.....	86
INTRODUCTION	87
METHOD.....	88
<i>Participants</i>	88
<i>Measures</i>	89
<i>Apparatus</i>	91
RESULTS.....	92
<i>Verbal Fluency Task</i>	93
<i>Digit Span Forward</i>	94
<i>Digit Span Backward</i>	96
<i>Results for the Three IVR Tasks Combined</i>	97
DISCUSSION	100
REFERENCES	104
TABLE 1.	107
TABLE 2.	108
TABLE 3.	109
TABLE 4.	110
TABLE 5.	111
TABLE 6.	112
TABLE 7.	113
TABLE 8.	114
TABLE 9.	115
GENERAL DISCUSSION.....	116
CONCLUSIONS	121
REFERENCES	123
APPENDICES	128
APPENDIX A.....	129
APPENDIX B.....	130
APPENDIX C.....	131
APPENDIX D.....	137
APPENDIX E.....	138
APPENDIX F.....	139
APPENDIX G.....	140
APPENDIX H.....	142
APPENDIX I.....	143
APPENDIX J.....	145

General Abstract

The main goal of this project was to enhance the use and usability of Interactive Voice Response Systems (IVR) for older people. The objective of study one was to examine older people's perceptions of the technology and identify the most common difficulties experienced by older people when interactive with IVR using focus groups. Twenty-six people aged 65 and older took part in the study. Data were analyzed using frequency and chi square analyses. The results revealed negative attitudes towards the technology. Long menus, frustration about not being able to reach an operator and absence of shortcuts were some of the most common difficulties reported by participants.

Study two examined the cognitive factors predicting successful IVR interaction in four commercially available IVR systems in a sample of 185 older adults. Linear regressions were performed on the data. Results indicated that working memory and auditory memory were the best predictors of successful IVR interaction.

Using the same sample of participants as study two, study three examined older adults' attitudes towards the four IVR systems in relation to their success in interacting with the technology. The study also evaluated the impact of gender on success and attitudes towards IVR. There was a significant positive correlation between success with IVR and favorable attitudes towards the technology. No gender differences emerged in both performance on IVR tasks and attitudes towards the technology.

Study four evaluated the feasibility of using a voice-activated IVR to administer and score three short neuropsychological tests using a sub sample of the original sample of 185 older adults involved in study two and three. One hundred and fifty eight participants took part in the study. Results showed high correlations between the IVR and clinician scoring

of the three tests. Nevertheless, a number of discrepancies and technical issues were discovered.

General Introduction

Definition of IVR systems

The acronym IVR stands for interactive voice response technology. Other commonly used names include automated telephone answering systems, voice response units and automated attendants. IVR is a telephony technology in which people use a touch-tone phone or verbal responses to interact with a database. IVR does not require human interaction over the telephone; instead the caller interacts with a computer connected with the phone system to either acquire or enter information into a database. For example, banks and credit card companies use IVR systems to provide customers with up-to-date account information or in the case of telephone surveys, IVR systems are used to gather information from users.

IVR systems were first introduced in the 1980s but their potential was not immediately recognized. The integration of computer technologies and telephone systems in the 1990s was the contributing factor to their increased popularity (Dulude, 2002). In the last 20 years, there has been a rapid growth in the function of voice processing technology from simple voice mail, in which messages can be stored and retrieved to complicated systems that direct incoming calls and provide access to a wide range of pre-recorded messages.

IVR applications

An important advantageous feature of IVR systems is that they are accessed through the telephone. The telephone is still the most widely adopted technology with 99.1 % of Canadian households owning and using a telephone (home or cellular) compared to other communication devices such as the computer (79.4 %). When older adults are concerned,

98.9% of them own and use a telephone compared to only 33.1% who own a home computer, and 26.6% of them using internet at home (Statistics Canada, 2009).

The economic advantage of IVR systems, compared to live operators, is the main reason for their increasing popularity (Lennig, Bielby, & Massicotte, 1995). IVR systems have been shown to increase productivity by allowing employees to focus on other tasks as opposed to answering telephone calls (Dulude, 2002; Forster & van Walraven, 2007b; Lennig et al., 1995). IVR systems are also available, around the clock, and this is beneficial to both companies and customers. Another important benefit of using IVR systems as opposed to live operators is the perception of increased privacy when communicating sensitive information (Bardone, Krahn, Goodman, & Searles, 2000; Lee, 2003; Nakagawa, 2000).

In recent years, there has been an increased interest in adopting IVR technology in the healthcare system and in medical and psychological research. More specifically, two major healthcare areas associated with better outcomes for patients and significant reduction in medical cost -- follow-up of patients after hospital discharge and preventive care, have successfully used IVR technology. For example, research has demonstrated the feasibility and utility of IVR systems in follow-up of patients with multiple medical problems after hospital discharge and thus reducing the amount of time needed for trained practitioners to call patients (Forster & van Walraven, 2007a), and significantly lowering readmission rates (Graham et al. 2012). IVR systems have also been used successfully in monitoring patients with chronic conditions, improving medication and treatment adherence, and relapse prevention. Examples of such systems include using IVR for oral anticoagulant management (Oake, van Walraven, Rodger, & Forster, 2009), adherence to statin treatment (Stacy,

Schwartz, Ershoff, & Shreve, 2009), reminding patients with chronic conditions to take or refill medications (Reidel, Tamblyn, Patel, & Huang, 2008), monitoring and relapse prevention of drinking problems (Rose, Skelly, Badger, Naylor, & Helzer, 2012 ; Tucker, Roth, Huang, Crawford, & Simpson, 2012), and smoking cessation (Regan, Reyen, Lockhart, Richards, & Rigotti, 2011), treatment and relapse prevention of chronic pain (Naylor, Keefe, Brigidi, Naud, & Helzer, 2008), monitoring of post-traumatic stress symptoms and substance use in veteran population (Possemato et al. 2012), and providing peer support and increasing medication adherence of patients with chronic heart failure (Heisler et al., 2007).

In preventive care, IVR-delivered messages have been shown to increase the frequency of cervical and breast cancer screening, (Corkrey, Parkinson, & Bates, 2005; Crawford et al., 2005; DeFrank et al., 2009), immunization for influenza in target populations (Crawford et al., 2005), screening for depression (Kim, Bracha, & Tipnis, 2007; Moore et al., 2006; Rush et al., 2006), screening for post-partum depression in low-income mothers (Kim et al., 2012), screening for dementia (Mundt, Geralt, & Moore, 2006; Mundt, Kinoshita, Hsu, Yesavage, & Greist, 2007; Sano et al., 2010; Smith, Tremont, & Ott, 2008), population screening for risky alcohol and drug use (Mundt, Bohn, King, & Hartley, 2002; Sinadinovic, Wennberg, & Berman, 2011), screening for binge eating (Bardone et al., 2000), and assessment of sexual activity (Schroder, Johnson, & Wiebe, 2007). The utility and feasibility of IVR systems also extend to increasing adherence to medical appointments and pre-appointment procedures (Morrow, Leirer, Carver, Tanke, & McNally, 1999) as well as providing a number of IVR adapted behavioural interventions. Examples of IVR delivered interventions include promoting weight loss in pre-diabetic population (Estabrooks & Smith-

Ray, 2008), administering guided self-help behavioural treatment for patients with obsessive compulsive disorder (Greist, Marks et al., 2002), and providing self-help interventions for patients with mild and moderate depression (Osgood-Hynes et al., 1998).

In medical and psychological research, IVR systems have been shown to not only be cost-effective but also more reliable tools compared to traditional methods of data gathering. There is evidence to suggest that IVR screening of participants for inclusion in clinical trials, monitoring participants' symptoms during clinical trials, and self-monitoring of the resolution of addictive behaviours such as drinking and smoking is more reliable compared to screening and monitoring of symptoms by clinicians (Anderson, Soderpalm Gordth, & Berglund, 2007; Greist, Mundt, & Kobak, 2002; Kobak, Greist, Jefferson, & Katzelnick, 1996; Toll, Cooney, McKee, O'Malley, & Cooney, 2008; Tucker, Foushee, & Black, 2008). The higher reliability of IVR systems in research settings has been associated with their objectivity as well as the increased perception of privacy by participants when communicating sensitive issues (Greist, Mundt et al., 2002).

In summary, it appears that IVR systems have the potential to save time and money for companies in providing customer service, aid practitioners in delivering healthcare services to patients, improve the quality of medical and psychological research, and are a feasible option to some in-person screening and assessment. However, there are some obstacles to using IVR systems, particularly in an older adult population.

IVR Systems and Older People

Despite the rapid adoption of IVR systems, research regarding public attitudes and usability of these systems is still scarce. In fact, in 1997 when Katz and colleagues conducted their survey of public attitudes of U.S. population towards IVR systems, they

found no academic literature on the topic. In contrast, web-based computer technologies have received considerably larger attention with numerous studies evaluating factors such as gender, race, class, age, education level and cognitive ability on the use of the technology (Birdi & Zapf, 1997; Dyck & Smither, 1999; Jay & Willis, 1992; Westerman, Davies, Glendon, Stammers, & Matthews, 1995). The results of Katz et al. (1997) survey of public attitudes regarding several telecommunication technologies such as the telephone, voice-mail and IVR in 912 Americans, showed that the single demographic variable most strongly related to public attitudes towards IVR was age. In the survey, older adults reported more negative perceptions of the technology. Surprisingly, education level, income and gender were not related to attitudes towards IVR, which was in contrast with what was previously found in the literature regarding technology in general and the use of the telephone technology in particular (Katz, 1994). The variable most strongly associated with positive attitudes towards these systems was the quality of people's most recent experience with the technology. Compared to younger adults, older adults also reported more unsatisfactory experiences with IVR; in fact only 55% of people 65 years of age and older reported that during their last encounter with an IVR they got through to the person they wanted to talk to and only 63% indicated that they had accomplished the goal that prompted their call. The survey also identified some general areas of dissatisfaction with IVR systems; these included difficulties in describing the problem for which they were calling to an IVR (70% of participants), having to listen to irrelevant options (70%), worries that the message they left with an IVR will not be received by anybody (52%), and encountering choices that led to nowhere and prevented them from achieving their goal (42%). The convenience of being able to receive a service at any time was identified as one area of satisfaction. However, a

large number of respondents felt that the companies derived much more benefits from using IVR than the costumers. In addition, only 20% of respondents felt that they were in control while using IVR versus 53% who disagreed with the notion (Katz, Aspden, & Reich, 1997).

These results are in line with the findings from another U.S. survey of 800 adults and their attitudes towards automated phone systems, which showed overall negative attitudes towards IVR and a tendency for stronger dislike of the technology by older respondents (Settle, Dillon, & Alreck, 1999). Older respondents also reported less frequent use of the technology. Once again, a number of participants reported frustration related to having to listen to long list of options unrelated to their call, which resulted in longer times needed to complete the call. More than a quarter of respondents indicated that the IVR systems they had encountered were not set up in a way that they ought to be. Based on the survey responses, the authors identified three main areas for improvement of IVR systems; these included considering the impact and the effect of the IVR systems on callers during the design phase, balancing the depth and the breath of the menus (fewer options at each level of the system), and having operators available to respond in person, if needed (Settle et al., 1999). The second recommendation appears to be of a special importance to older users as numerous menu options place greater memory demands on callers.

The relationship between age and success in using IVR was tested experimentally in 22 community dwelling older women and 22 female university students (Dulude, 2002). In this study, participants were asked to perform real-life IVR tasks using six different commercially available IVR systems. Participants were also asked to complete a short usability questionnaire in order to evaluate their impressions and attitudes regarding the six IVR systems used in the study. The results showed that increasing age significantly reduced

participants' performance on the IVR tasks: older women completed fewer tasks than younger women. However, a subset of older women performed just as well as younger participants did, and the distribution of older participants success with the IVR systems was bimodal. Not surprisingly, participants did better on the tasks that required fewer choices (levels) to complete. The only voice-activated IVR in the study was an exception in that respect, as younger adults did better on the task regardless of the number of choices required. Additionally, there was a large discrepancy between the ratings of older and younger women for that system. Younger participants gave significantly higher ratings than older adults. The authors hypothesized that it may be that voice-activated systems follow different rules than keypad-based IVR. In Dulude's study, the hypothesis that older women gave lower usability ratings to the IVR was supported for only two out of six IVR systems. The most commonly reported difficulties by older women in using the six IVR included confusing choices of instruction, voices speaking too quickly, lengthy introductions, menus and items, problems when entering verbal and keystroke data, callers ignoring certain prompts, confusion about the task, no means of recovery from mistakes, and use of jargon. Additionally, analyses of the errors committed by participants during their interaction with the systems, revealed that older and younger participants encountered the same difficulties. However, younger participants were more successful in recovering from errors compared to their older counterparts. These results were indicative of interplay between interface design flaws and age. Factors related to age might have included cognitive factors such as attention, memory, and reaction time. For example, in this study, older adults were seven times more likely to forget that one of the tasks required them to obtain flight information about "tomorrow" and not "today". Additional factors that might also have played a role in older adults' success in

interacting with the IVR systems included sensory and motor problems, lack of familiarity with new technologies and jargon (Dulude, 2002).

To date, only two studies have attempted to examine the relationship between cognitive ability and successful interaction with IVR. The first study, an unpublished manuscript by (Dulude, 2000), found no association between participants' performance on a working memory task and their IVR performance. It is conceivable that the lack of relationship between working memory and performance on IVR was due to the small sample size of the study (22 older and 22 younger participants) and the use of only one working memory task. Another study, using a larger sample of 196 participants (73 men and 123 women) between 18 and 91 years of age, examined the effects of age, speech rate and cognition on performance on two simulated IVR systems (Sharit, Czaja, Nair, & Lee, 2003). Once again, results revealed higher success rate on IVR tasks for the younger participants. Although the authors included a battery of 20 cognitive tests in their study, they used only two of them (a test of attention and a test of working memory) in their final analyses. Results showed a significant relationship between working memory and performance on IVR tasks (Sharit et al., 2003). However the authors did not evaluate and/or report on the association between the other cognitive domains measured in their study and IVR performance.

Findings from numerous studies on aging and cognitive function show that normal aging, particularly after the age of 60, is associated with the decline of a number of cognitive functions such as working memory, executive functions and processing speed (Anstey & Low, 2004; Gazzaley & D'Esposito, 2007; Gazzaley, Sheridan, Cooney, & D'Esposito, 2007; Park, O'Connell, & Thomson, 2003; Treitz, Heyder, & Daum, 2007; Verhaeghen & Cerella, 2002). Another important finding from the research on aging and cognitive function is the

greater variability in cognitive abilities of older people compared to younger people, suggesting that older people differ significantly with respect to the rate at which their cognitive functions change over time (Anstey, 2004; Glisky, Rubin, & Davidson, 2001; Rosen et al., 2002; Treitz et al., 2007). This is also consistent with the findings from the Dulude's (2000) study, which found that the IVR performance of older participants had a bimodal distribution; half of the older women performed very poorly, a third performed extremely well, and the performance of a small proportion of older participants fell in the middle.

To summarize, there is evidence supporting the notion that normal aging is accompanied by changes in cognitive function and that these changes may be related to older adults' difficulties in interacting with IVR systems. Given the rapid adoption of IVR systems in providing a number of services from business to healthcare, seniors appear to be at a disadvantage in benefiting from IVR-provided services. It appears that a more adaptive approach to developing IVR systems is needed in order to address the cognitive changes that a number of older people face in order to allow them to have the same access to IVR-provided services as younger people.

The Present Study

The present study aimed to improve the use and usability of IVR systems for older adults. The goal of study one was to examine the most common difficulties that this population encounters in their interactions with IVR systems and evaluate seniors' attitudes towards the technology using focus groups. Two focus groups were conducted with 26 participants aged 65 years and older. Participants' responses during the two focus groups were audio recorded and subsequently transcribed and analysed. Participants involved in the

focus groups were not included in the other studies. Study 2 examined the cognitive and memory factors related to older adults' difficulties with these automated systems by evaluating the extent to which participants' cognitive abilities predicted their performance on four publicly available IVR systems. There were 185 participants over the age of 65 who took part in this study. In addition to the four IVR systems, participants also completed the Wechsler Adult Intelligence Scale fourth edition and the Wechsler Memory Scale fourth edition (older adult battery). Study three evaluated participants' attitudes towards IVR systems in relation to their success or failure in interacting with these four IVR systems in the same sample of 185 participants involved in study two. Attitudes were measured using a questionnaire, which gathered information regarding the IVRs' ease of use and perceptions about the automated systems. The study also evaluated the impact of demographic variables such as gender, education and full-scale IQ on participants' attitudes towards the systems. Study four was a proof-of-principle study, which examined the feasibility of using an IVR for the administration and scoring of neuropsychological tests. The study adapted three commonly used neuropsychological tests (verbal fluency and digit span forward and backward) for an IVR administration and scoring and evaluated the validity and reliability of the IVR tests to an in-person administration and scoring.

I was directly involved in all stages of these projects: project planning, design of the IVR-COG application, data collection, data analyses, writing of manuscripts.

The present study was a collaboration between TelAsk Technologies and a team of researchers at the University of Ottawa. The target population of the study was adults aged 65 and older. The study and all corresponding documentation was reviewed and approved by

the Social Sciences and Humanities Research Ethics Board (SSH REB) at the University of Ottawa on July 29th 2008.

Originality of the Project

To date, only two studies have attempted to examine the relationship between cognitive abilities and IVR performance; both of these studies used a very limited number of cognitive measures. The present study aimed to provide a more comprehensive evaluation of the relationship between cognition and ability to cope with IVR in older adult population using two batteries of cognitive tests, which assess a wide range of cognitive abilities.

To date, a number of applications have employed IVR systems to administer cognitive tests over the telephone, but very few of them have developed a system that both administers and scores the evaluation online. In addition, none of the previously used applications have used verbal responding as opposed to key-activated IVR when the automated system was used to score the evaluation. Voice-activated systems face unique challenges in voice recognition; thus, the results of this study shed light on some of the benefits in using voice-activated IVR for the purpose of neuropsychological evaluation and also highlighted some of the limitations.

Study one entitled “Improving older adults’ experience with interactive voice response systems” evaluated older adults’ attitudes towards IVR systems and examined the most common difficulties experienced by older adults when interacting with the technology using two focus groups. Participants were also asked to provide feedback on the voice-activated IVR system we developed in collaboration with TelAsk Technologies. The system (named IVR-COG) administered and scored three short neuropsychological tests. In study four, we presented the results regarding the feasibility of using voice-activated system such as the IVR-COG for the purpose of administering and scoring neuropsychological evaluations. The study was published in the journal of Telemedicine and E-Health in 2011. I was involved in the design of the experiment, recruitment of participants, conducting the focus groups, analyzing the data and writing of the manuscript. Halina Bruce was involved in coordinating the project and reviewing versions of the manuscript. Dr. Michele Gagnon was involved in the role of a consulting neuropsychologist during the design stages of the project. Vincent Talbot from TelAsk Technologies contributed in the conceptualization, design and implementation of the IVR-COG. He took part in the focus groups and reviewed versions of the manuscript. Dr. Claude Messier supervised the project and was the holder of the NSERC grant that funded the research.

Miller, D. I., Bruce, H., Gagnon, M., Talbot, V., & Messier, C. (2011). Improving older adults’ experience with interactive voice response systems. *Telemedicine and E-health, 17*(6), 452- 455.

Running head: OLDER ADULTS AND IVR SYSTEMS

Improving older adults' experience with interactive voice response systems

Delyana Ivanova Miller, B.A., Halina Bruce, B.A., Michèle Gagnon, PhD, Vincent Talbot,

BSc. and Claude Messier, PhD

Abstract

Interactive voice response systems (IVR) use computer-based voice recognition and software algorithms to conduct human/computer interactions. In recent years, there has been a proliferation of IVR applications in business and healthcare. The available evidence suggests that older people have negative attitudes towards IVR and experience significant difficulties using these systems. **Objective:** The goal of this project was to identify areas of difficulties in IVR use by older people and suggest strategies for improvement. **Materials and Methods:** We used two focus groups to examine older people's perceptions of IVR systems and the most common difficulties experienced in interacting with these systems. We also recorded suggestions for improvement of IVR. **Results:** Frequency and chi square analyses were performed on the focus groups' data. Some of the difficulties reported by participants in this study were congruent with previous findings, but we also uncovered some additional problems, such as frustration for not being able to reach an operator, being asked to wait too long on hold, being unable to recover from mistakes and an absence of shortcuts in the systems. In addition, a significant number of participants indicated that they prefer a system that adjusts to them automatically as opposed to a system that allows for adjustment. **Conclusion:** Generally, our findings suggest that the poor acceptability of IVR systems by older people could be improved by designing IVR algorithms that detect difficulties during an ongoing IVR exchange and direct people to different algorithms adapted for each person.

Introduction

Interactive Voice Response (IVR) is a telephone technology that uses a touch-tone telephone or verbal responses to interact with a computer program (Rottenberger, 1991). A number of factors explain the proliferation of IVR systems. The telephone is still the most widely adopted technology. In Canada, 99.1 % of households own and use a telephone (home or cellular) compared to other communication devices such as the computer (79.4 %). Within the older population, 98.9% own and use a telephone compared to 33.1% for home computers, and only 26.6% of older adults access the internet at home. Another obvious advantage is that IVR systems are less costly than live operators (Lennig, Bielby, & Massicotte, 1995) and are continuously available. Finally, IVR systems give callers the impression of increased privacy compared to live person interaction particularly when communicating sensitive information (Bardone, Krahn, Goodman, & Searles, 2000; Lee, 2003; Nakagawa, 2000).

The use of IVR technology for healthcare applications and for medical and psychological research is growing. Some of the healthcare applications of IVR systems include follow-up of patients after hospital discharge (Forster & van Walraven, 2007; Lee, 2003), peer support to patients with chronic heart failure (Heisler et al., 2007), weight loss promotion in pre-diabetic patients (Estabrooks & Smith-Ray, 2008), increase of medication adherence (Oake, 2009; Stacy, 2009), and post-operative surveillance and care (Forster, Boyle, Shojania, Feasby, van Walraven, 2009; Sherrard, 2007).

Acceptability is an important issue for IVR use. To date, only one study has examined people's attitudes towards IVR systems. This study revealed that the demographic variable most strongly related to negative attitudes towards IVR systems was age; older participants had more negative attitudes and more unsatisfactory experiences with IVR systems

compared to younger people (Katz, Aspden, & Reich, 1997). The only study that has investigated the relationship between age and IVR performance, as measured by the number of completed tasks on six different IVR systems (Dulude, 2002), found that older age was associated with fewer completed tasks. The factors underlying older people's attitudes and difficulties in interacting with these systems are still not understood, mainly because of the lack of research in the area.

In the present report, we present the results of focus groups, which were designed to identify problematic issues in IVR interaction for older people. We examined older people's perceptions of IVR systems and the most common difficulties experienced in interacting with these systems. We also recorded suggestions for improvement of IVR.

Materials and Methods

Participants were 65 years and older with good proficiency in English. Participants of the focus groups were recruited through advertisements. To recruit people from lower socioeconomic status, advertisements were also placed in subsidized housing. For the purpose of this study, we developed an IVR application called IVR-Cognition (IVR-COG), which was used as an example of an IVR system for two focus groups. The IVR-COG consisted of two cognitive tasks: the Naming Task (Teng & Chui, 1987), in which participants were asked to name as many fruits as they could within 1 min, and the Digit Span forward and backward task, in which participants were asked to repeat strings of single digit numbers of increasing length. Participants were told that one of the goals of the study was to find out whether their performance on these three tasks could eventually be used to adjust the IVR system to each individual. For example, for people with poorer memory, the IVR system could present a fewer number of alternatives for each question. This strategy was not implemented but was used to probe whether people would submit to these three

short tests in order to get a “personalized” IVR interaction. Participants could also adjust the volume of the IVR voice and choose from three different paces of speech delivery.

A number of open-ended questions probed participants about their previous experiences and attitudes towards IVR systems. In addition, there were a number of more specific questions about the ease of use of the IVR-COG system. Participants were invited to share ideas about characteristics that could improve their experience with that specific system and with IVR systems in general. This study was approved by an Institutional Review Board.

Procedure

Participants met the first time as two separate groups based on education level achieved, whereas all participants met together for a second focus group meeting. The lower education group consisted of 10 participants; the higher education group consisted of 16 participants. All participants who took part in the first focus group were invited to participate in the second focus group (18 participants of the original sample of 26 attended the second focus group). General demographic information (age, education and employment history) was gathered to ensure the diversity of the sample. Participants were also informed that they would receive \$100 for their participation in the study.

A professional moderator who had previous experience leading focus groups for the purpose of developing IVR systems led the two groups. The moderator was responsible for maximizing participants’ involvement and clarifying comments. Using participants’ comments and suggestions from the first to the second focus group, spiral development methodology with multiple prototyping (Boehm, 1988) was used to improve the IVR-COG. The improved version of the IVR-COG was presented to participants in the second focus

group. In addition, before participants came to the second focus group, they were asked to call and try out the IVR-COG from home, to be able to more fully appreciate the experience with the system and give us their opinion on it. The two focus groups were audio recorded.

Results

Demographics

The mean age of participants in the lower education focus group was 70.9 years ($SD=5.0$), with age range 66-83 years. The mean age of participants in the higher education focus group was 71.2 years ($SD=5.8$), with age range 65-84 years. The lower education focus group consisted of 4 men and 6 women. The higher education focus group consisted of 6 men and 10 women. The mean education level of the lower education focus group was 11.1 years ($SD=1.5$), whereas it was 16 years ($SD=0.8$) for the higher education group. Two of the focus group participants were of Asian descent and the rest were Caucasian.

First Focus Group Analyses

Most common difficulties with IVR use

Since there were no significant differences between the two educational groups, chi-square analyses were performed on the combined results of these groups. Most participants (92 %) indicated that they had interacted with an IVR system in the last six months ($\chi^2=18.6$, $df=1$, $p < 0.5$). Several themes emerged when participants were asked about their experiences with these IVR systems. Participants indicated that the most common difficulties they experienced with IVR systems were confusing instructions, waiting too long to reach an operator, being cut off by the system for not responding fast enough (or making a mistake), and irrelevant instructions (at least from the caller's perspective). Other complaints included voices speaking too quickly, too many options to remember and choose from,

absence of shortcuts in the system (which at times made the interaction tedious), frustration about being unable to reach an operator, lengthy IVR interaction, and being asked to provide the same information more than once. One of the few aspects that participants reported liking about the IVR systems was that IVR systems are “patient” with them (they “don’t mind” repeating many times the same things). Despite this single advantage, most people reported that they dislike IVR systems and avoid using them whenever possible.

Vocal characteristic preferences

Individuals were presented with six different voices (three male and three female voices) and subsequent analyses were performed to reveal participants’ preferences. Most participants reported that it was not important which voice they were interacting with, as long as the voice was clear, with 62% of participants reporting no preference for age or gender. Significantly more participants preferred interacting with a “persona” with a slightly artificial but clear voice as opposed to a naturally sounding voice that might not be as clear (92% of participants ($\chi^2=18.6, 01, df=1, p < 0.5$). All participants reported that they preferred to interact with an IVR persona that was efficient and direct as opposed to a persona who was chatty and friendly. When participants were asked how they felt about less formal language and familiarity over the phone via an automated system, all participants reported that they would not like it.

System interaction preferences

All participants in both the higher and the lower education focus groups indicated that they would prefer to be told at the beginning of the interaction that they were speaking to a computer and not a real person. They reported that otherwise there was a possibility to mistake the computer for a real person and this would make them feel uncomfortable (silly)

upon realizing their mistake. We observed in other studies that not realizing or forgetting that one is speaking to a computer during the interaction is a common occurrence in older people. All participants in the lower education focus group also reported that phrases such as “you are interacting with an automated system” would be confusing; they would rather be told that they are “talking to” a computer.

Participants were asked if they would prefer interacting over the phone with a grumpy real person or a friendly IVR system: 42% reported that they prefer to interact with a friendly IVR system, 19% preferred the interaction with the grumpy real person over a friendly IVR system, and 39% of participants reported no preferences ($\chi^2 = 8.96$, $df = 2$, $p < 0.05$). Significantly larger number of participants (65%) reported that they would rather communicate sensitive issues to a real person than to an IVR system ($\chi^2 = 13.01$, $df = 2$, $p < 0.05$).

Participants were asked what was the maximum amount of time they would wait for a real person to return their phone call assuming they were not willing to interact with a computer system: 81% of participants indicated that they would wait for 24 hours, and 1% of participants reported that they would expect their call to be returned the same day.

Participants were presented with a situation in which a patient who underwent surgery is discharged from the hospital. Participants were asked their preference regarding follow-up calls from the hospital assuming that there were no resources available to employ a real person for the task. Most participants (94%) reported that they would prefer to be called by an IVR system following hospital discharge if the alternative was to not be called at all.

When participants were asked whether they preferred to be called by the computer or if they would rather initiate the call themselves, 85% of participants reported that they would rather

initiate the call, less than 1% reported that they would rather have the system call them, and less than 15% of participants reported that it makes no difference to them.

When participants were asked to describe a perfect IVR system, themes that emerged were: friendly, polite, with short, clear and unambiguous instructions. The majority of participants reported that the opportunity to adjust the volume and the speed of the conversation was very useful, (92% ; $\chi^2 = 18.6$, $df = 1$, $p < 0.5$). Many participants (31%) reported that they disliked motivational phrases such as “you did well” and “good job” at the end of the IVR tasks because they sounded patronizing to them. We asked participants whether, in the case of a mistake, they would rather have the instructions repeated exactly the same way or with slight variations. Most participants (77%) reported that they preferred if error messages or instructions remained the same each time they were repeated. Thirty-eight percent of participants indicated that an option to pause the IVR task (i.e., for a bathroom break or to answer the door) would be very useful, so that they would not have to start over from the beginning. Length of the interaction was also a problem: 30% of participants indicated that by the time they reached the third task they were tired. Finally, 54% of people reported that they would be willing to do all three tasks if this would make their interaction with the system easier.

Second Focus Group Analyses

All participants indicated that they would like the option to go back and readjust the volume and pace of the conversation, in case they change their mind about their previous selection. Two people who were wearing hearing aids reported that they would have preferred to interact with a male voice, because it was easier for them to hear male voices. The preference of a male voice as opposed to female one by people with hearing difficulties

may be due to lower auditory frequencies in male voices. An IVR system that allows callers to choose the gender of the voice for the IVR interaction may address this issue. When participants were asked whether they would agree to be tested by the IVR system with the IVR-COG every time they called, 44% reported that they would like to take the test each time, 25% reported that they would like to be given the option to take or not take the test every time, and 31% reported no preference. All participants indicated that they would prefer if the system would adjust itself to a slower pace if they required instructions to be repeated frequently, as opposed to be asked by the system if they would like to slow down the pace. The rationale for the automatic adjustment was that participants would feel embarrassed by their difficulties and might not ask to slow the pace. All participants also reported that they prefer to interact with the system verbally as opposed to using the telephone keypad. Interacting via a handheld telephone keypad sometimes results in missing the beginning of messages because of the delay between listening from the receiver, looking at the keyboard, punching numbers, and returning the receiver to the ear.

Discussion

The current study examined the most common difficulties encountered by seniors when interacting with IVR systems and recorded proposed solutions on how to improve the use and usability of these systems for older adults. Older people in our sample reported that they use IVR systems, but their overall experience with these systems was negative. Although there were a few redeeming attributes of IVR systems (such as not getting impatient after numerous repetitions of instructions), most participants tried to avoid interacting with IVR systems. In addition, several participants reported that they would prefer an interaction with a grumpy real person to the alternative friendly IVR system, indicating that a number of seniors would opt out for a live person interaction no matter how unpleasant the live person

is. One surprising finding, which went against previous research data, was that many participants reported preference of communicating sensitive information to a live person rather than to an IVR system.

Some of the difficulties reported by participants in this study were congruent with previous findings, such as long menus, confusing instructions and voices speaking too quickly (Dulude, 2002). We also uncovered some additional problems; these included frustration at not being able to reach an operator, being asked to wait too long on hold, being unable to recover from mistakes and an absence of shortcuts in the systems.

The findings regarding seniors' difficulties during an interaction with IVR systems and their negative attitudes towards these systems were also in line with one of the leading theoretical approaches in social psychology concerned with the factors related to the acceptance of technology by consumers -- the Technology Acceptance Model (Davis, 1989). According to this model, the two important factors that determine one's attitudes, acceptance, and use of a specific technology are the perceived ease of use of the technology and its usefulness. We found that older adults experienced a number of difficulties while using IVR systems and that their attitudes towards the technology were generally negative. However, when seniors were presented with a possible healthcare application of an IVR system that would provide a service, which would not be available without such a system, they reported that they would be willing to use the IVR.

It was clear that most older adults do not like to be reminded that they may be experiencing age-related cognitive decline and that having difficulties with an IVR system is taken as a reminder of this decline. Within this context, participants preferred a self-adjusting IVR system, because it does not explicitly reveal their lack of competence.

Additionally, participants in this study resented the words of encouragements proffered by the system and perceived these as condescending. Generally, our findings suggest that the poor acceptability of IVR systems by older people could be improved by designing IVR algorithms that detect difficulties during an ongoing IVR exchange and direct people to different algorithms adapted for each person.

Acknowledgements: This research was funded through a grant from the Natural Sciences and Engineering Research Council of Canada and by TelAsk Technologies.

Disclosure Statements: Vincent Talbot is a consultant for TelAsk Technologies. TelAsk Technologies funded part of this project. No other conflict of interest for C. Messier, D.I. Miller, H. Bruce and M. Gagnon.

References

- Bardone, A. M., Krahn, D. D., Goodman, B. M., & Searles, J. S. (2000). Using interactive voice response technology and timeline follow-back methodology in studying binge eating and drinking behavior: different answers to different forms of the same question? *Addict Behav*, 25(1), 1-11.
- Boehm, B. W. (1988). A spiral model of software development and enhancement. *Computer*(May), 61-72.
- Davis, F. D., Bagozzi, R. P., Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *JSTOR: Management Science*, 35(8), 982-1003.
- Dulude, L. (2002). Automated telephone answering systems and aging. *Behaviour & Information Technology*, 21(3), 171-184.
- Estabrooks, P. A., & Smith-Ray, R. L. (2008). Piloting a behavioral intervention delivered through interactive voice response telephone messages to promote weight loss in a pre-diabetic population. *Patient Educ Couns*, 72(1), 34-41.
- Forster, A. J., Boyle, L., Shojania, K. G., Feasby, T. E., van Walraven, C. (2009). Identifying patients with post-discharge care problems using an interactive voice response system. *JGIM*, 24(4), 520-525.
- Forster, A. J., & van Walraven, C. (2007). Using an interactive voice response system to improve patient safety following hospital discharge. *Journal of Evaluation in Clinical Practice*, 13, 346-351.
- Heisler, M., Halasyamani, L., Resnicow, K., Neaton, M., Shanahan, J., Brown, S., et al. (2007). "I am not alone": the feasibility and acceptability of interactive voice response-facilitated telephone peer support among older adults with heart failure. *Congest Heart Fail*, 13(3), 149-157.
- Katz, M. J., Aspden, P., & Reich, W. A. (1997). Public attitudes toward voice-based electronic messaging technologies in the United States: A national survey of opinions about voice response units and telephone answering machines. *Behaviour & Information Technology*, 16(3), 125-144.
- Lee, H., Friedman, M. E., Cukor, P., Ahem, D. (2003). Interactive voice response system (IVR) in health care services. *Nurs Outlook*, 51(6), 277-283.
- Lennig, M., Bielby, G., & Massicotte, J. (1995). Directory assistance automation in Bell Canada: Trail results. *Speech Communication*, 17, 227-234.

- Nakagawa, A., Marks, I. M., P, J.M., Bachofen, M., Bear, L., Dottl, S.L., Greist, J.H. (2000). Treatment of obsessive-compulsive disorder guided by manual and computer-conducted telephone interview. *Journal of Telemedicine and Telecare*, 6, 22-26.
- Oake, N., van Walraven, C., Rodger, M. A., Forster, A. J. . (2009). Effect of an interactive voice response system on oral anticoagulant management. *CMAJ*, 180(9), 927-1004.
- Rottenberger, K. (1991). The next voice you hear... *Sales and Marketing Management*, 143(8), 50-55.
- Sherrard, H. (2007). Providing innovative clinical services with a patient focus. *The Ottawa Heart Institute Alumni Bulletin*, 16(1), 3.
- Stacy, J. N., Schwartz, S. M., Ershoff, D., Shreve, M. S. (2009). Incorporating tailored interactive patient solutions using interactive voice response technology to improve statin adherence: results of randomized clinical trial in managed care setting. *Population Health Management*, 12(5), 241-254.
- Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry*, 48(8), 314-318.

Study two entitled “Predictors of successful communication with interactive voice response systems in older people” examined the cognitive factors involved in seniors’ ability to interact successfully with IVR systems. The study included a different sample of participants than the one used in the focus groups and its objective was to expand our understanding on the important factors involved in IVR interaction in older people. The study was published in the *Journal of Gerontology* in 2012. I was involved in the design of the experiment, recruitment of participants, testing of participants, analyzing the data and writing of the manuscript. Dr. Michele Gagnon was involved in the role of a consulting neuropsychologist during the design stages of the project. She also provided training in neuropsychological test administration of the WAIS-IV and WMS-IV to the research assistants involved in the study. Vincent Talbot from TelAsk Technologies reviewed versions of the manuscript and provided feedback. Dr. Claude Messier supervised the project and was the holder of the NSERC grant that funded the research.

Miller, D., Gagnon, M., Talbot, V., & Messier, C. (2012). Predictors of successful communication with interactive voice response systems in older adults. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, DOI: 10.1093/geronb/gbs092.

Running head: PREDICTORS OF SUCCESSFUL IVR COMMUNICATION

Predictors of Successful Communication With interactive Voice Response Systems in Older
People

Delyana Miller, Michèle Gagnon, Vincent Talbot and Claude Messier

Abstract

Interactive voice response systems (IVR) are computer programs that can interact with people to provide a number of services from business to health care. However, surveys examining people's attitudes towards these systems have consistently found that people in general and older people in particular strongly dislike these systems. **Objective:** We wanted to determine the memory and cognitive abilities that predict successful IVR interactions for older people. **Method:** We compared the performance of 185 older adults (aged 65 and older) on normed cognitive tests (the Wechsler Adult Intelligence Scale fourth edition and the Wechsler Memory Scale fourth edition) with their performance on four real-life IVR systems that included fact-finding at governmental agencies and plane ticket reservation. **Results:** The results indicated that adults over the age of 65 experience significant difficulties in interacting with IVR systems. A large number of participants (20.5%) could not complete any of the tasks; they tended to be older and had the lowest full-scale IQ. However, there was little difference between the age of participants who completed 1, 2, 3, or 4 tasks. Rather, auditory memory and working memory were the best overall predictors for success in IVR tasks. **Discussion:** The impact of poorer auditory and working memory is compounded by programming practices that increase the demand on these abilities and create unnecessary difficulties. Successful use of IVR systems could eventually complement in-person health services.

Introduction

Interactive Voice Response technology (IVR) is a telephone technology in which one uses a touch-tone telephone or verbal responses to interact with a computer.

The telephone is still the most widely adopted communication technology with 99.1 % of Canadian households owning and using a telephone (home or cellular). Virtually all older adults own and use a telephone (98.9%) whereas only 33.1% own a home computer, and 26.6% of them use the internet at home (Statistics Canada, 2009).

There has been a growing interest in adopting IVR technology in the healthcare system and in medical and psychological research. IVR systems can provide patient follow-up after hospital discharge and reduce the amount of time needed for trained practitioners to call patients. In one example, IVR was used to triage patient calls after their discharge following cardiac surgery; the IVR referred those who needed medical attention to a nurse (Forster & van Walraven, 2007; McPhail et al., 2010). IVR systems have also been used successfully to provide peer support to patients with chronic heart failure and improve self-care and medication adherence (Granger & Bosworth, 2011; Heisler et al., 2007). Preventive care is another healthcare domain in which IVR systems have been shown to be effective in increasing health behaviours. Several studies have shown the utility and feasibility of IVR systems in increasing the frequency of preventive care behaviours such as screening for cervical cancer, breast cancer, immunization for influenza in target populations, screening for post-partum depression, and adherence to hospital appointments (Corkrey, Parkinson, & Bates, 2005; Corkrey, Parkinson, Bates, Green, & Htun, 2005; Crawford et al., 2005; Fiscella et al., 2011; Hasvold & Wootton, 2011; Kim et al., 2012).

A number of screening measures have been adapted for IVR administration, such as screening for depression (Kim, Bracha, & Tipnis, 2007; Moore et al., 2006; Mundt, Geralt,

& Moore, 2006; Rush et al., 2006), alcohol use (Mundt, Bohn, King, & Hartley, 2002), binge eating (Bardone, Krahn, Goodman, & Searles, 2000), and sexual activity (Schroder, Johnson, & Wiebe, 2007). All of these measures have been shown to have a satisfactory degree of agreement with an in-person administration. In summary, it appears that IVR systems have the potential to increase patient accessibility to healthcare services, have proven reliable in medical and psychological research, and are a feasible alternative to a number of in-person screening and assessment procedures. However, most people, particularly older people, strongly dislike using IVR systems and experience significant difficulties when interacting with the technology (Katz, Aspden, & Reich, 1997; Miller, Bruce, Gagnon, Talbot, & Messier, 2011).

One study examined the contribution of cognitive factors to the performance of two IVR tasks: a banking IVR application to perform various banking tasks and an IVR application to request information or actions from an utility company (Sharit, Czaja, Nair, & Lee, 2003). Cognition was measured using Trails B and the Alphabet Test. The authors found an association between the results of these tests and IVR performance. Unfortunately, age and education were not controlled for in the regression analyses leading to this conclusion and the range of cognitive tests used was also very small. Another study examined the relationship between age and IVR performance as measured by the number of accomplished tasks on six different IVR systems (Dulude, 2002). The results showed that age was associated with poorer performance on IVR tasks. In Dulude's (2002) study, the most commonly reported difficulties of older participants were confusing choices of instruction, voices speaking too quickly, long introductions, menus and items, problems when entering verbal and keystroke data, callers ignoring certain prompts, confusion about

the task, no means of recovery from mistakes, and use of jargon. Similar conclusions were reached in a focus group analysis of attitudes toward IVR in older people (Miller et al., 2011). Previous research has shown that adoption of computer technology and use of the Web in older people was associated with better cognitive abilities and less computer anxiety (Czaja et al., 2006) and in general, older people fared poorly on web-based health information seeking (Czaja, Sharit, & Nair, 2008). It appears likely that older peoples' failure to successfully interact with IVR systems may be related to cognitive factors such as attention, memory, and reaction time. Additional factors that may also play a role in the ability to interact with an IVR system include sensory and motor problems, rate of delivery of spoken messages, lack of familiarity with new technologies and use of technical jargon (Sharit et al., 2003). Finally, the design of IVR systems typically does not take into consideration the physical and cognitive changes associated with normal aging.

Normal aging is associated with both a decline of a number of cognitive functions as well as a greater variability consistent with the known variable rate of cognitive aging (Park, O'Connell, & Thomson, 2003; Treitz, Heyder, & Daum, 2007; Verhaeghen & Cerella, 2002). Because the IVR interactions involve auditory/verbal exchanges and that each answer in the IVR interaction requires "online" information manipulations, we hypothesize that verbal/auditory memory and working memory will be the most determinant functions for success in an IVR interaction. We were also curious to see if other cognitive attributes such as perceptual reasoning abilities, processing speed abilities, and verbal comprehension abilities could also contribute to IVR successful interactions. To that effect, after completing real-life IVR applications, participants completed the Wechsler Adult Intelligence Scale (WAIS) and the Wechsler Memory Scale (WMS). Thus, the goal of the present study was to

measure cognitive and particularly memory functions using standardized tests in people aged 65 and older and examine the relationship between estimates of cognitive functions and older people's ability to successfully complete real-life IVR applications.

Method

Participants

The present study and all corresponding documentation was reviewed and approved by the Research Ethics Board of the University of Ottawa. One hundred and eighty-five (120 females: 65%) community dwelling people between 65 and 92 years of age (mean = 73.32 years, standard deviation [SD] = 6.44) were recruited from diverse socio-economic backgrounds, using advertisements in two free magazines for seniors and flyers, in community centers and subsidized housing buildings. Participants' education ranged from 7 to 22 years (mean = 13.88 years, SD = 2.90); 3% of participants had grade 8 or less, 14% had between grade 9 and grade 12, 23% had a high school diploma, 27% had some college or university, and 33% had a bachelor's, graduate, or professional degree. The only exclusion criteria were age younger than 65 and lack of proficiency in English. Participants were compensated \$100. The health questionnaire gathered self-reported demographic and health information. Eighty-eight percent were Caucasian. Ten percent reported being diabetic, 1.1% reported having had a haemorrhagic stroke, 1.1% had been treated for a brain tumour, 1.6% reported another unspecified brain disease, and 0.5% had chronic hepatitis. Two percent reported currently seeing a psychiatrist, and 9.2% were currently being treated for depression. Thirty-five percent of the sample reported experiencing memory problems. Compared to the WAIS and WMS normative groups, we had a slightly younger, more highly educated sample of common community-based volunteer cohorts. Using Cohen's d values,

there was only a small difference between our sample and the normative data on three tests (Vocabulary, Symbol Search and Information), whereas the difference on Coding was moderate.

Measures

WAIS fourth edition

The WAIS fourth edition (WAIS-IV) comprises of 10 core sub-tests and 5 optional sub-tests, measuring a number of cognitive functions. The test provides four composite scores on the Verbal Comprehension scale, Perceptual Reasoning scale, Working Memory scale and the Processing Speed scale as well as a global index of intellectual functioning, the full-scale IQ (FSIQ). The WAIS-IV was normed on a population aged 16-90 years old (Wechsler, 2008). On average, it takes 67 min to complete the 10 core sub-tests.

WMS fourth edition

The WMS fourth edition (WMS-IV) older adult battery (for people 65-90 years old) consists of five subtests, three with both immediate and delayed conditions. The battery produces four index scores: Auditory Memory Index, Visual Memory Index, Immediate Memory Index and Delayed Memory Index. The WMS-IV is normed on a sample of people aged 16-90 years (Wechsler, 2009). The battery was co-normed with WAIS-IV. The test takes about an hour to complete.

IVR tasks

The first two systems chosen were governmental IVR systems: Statistics Canada and Service Canada. Statistics Canada is the source of information on various aspects of Canada, many derived from census data. Service Canada is the central agency that provides information on all government programs including old age pensions. The last two were the

IVR systems of United Airlines and Air Canada that customers can use to access flight information and reserve tickets. The United Airlines system used verbal responses from callers, while the other three systems require callers to use the telephone keypad to enter their responses.

Apparatus

Participants' interaction over the phone with the four automated systems was recorded in order to facilitate scoring. A touch-tone phone (MITEL 5212) was used to call the IVR systems. A Sony MP3 IC recorder (ICD-UX7 1F/UX81F) was attached to the phone line. Sony stereo headphones (MDR-XD200) were attached to the recorder. Thus, the examiners were able to listen to participants' interaction with the IVR as the interaction was unfolding.

Procedure

Upon arriving at the memory laboratory at the University of Ottawa, self-reported information on participants' health and memory status was obtained using a health questionnaire. Next, participants completed four IVR tasks over the phone. The tasks asked participants to call an IVR system and obtain specific information. Participants were instructed to use only the automated menus and not use the option that allowed them to speak to an operator. Instructions on the four different tasks were both verbally presented to participants and a list of these instructions was placed in front of them while they were completing the task. Participants were allowed to take notes while completing the tasks but most participants did not.

The instructions for task 1 required participants to call Service Canada and obtain information on what one needs in order to qualify for an old age security pension. The

instructions for task 2 asked participants to call Statistics Canada and obtain the latest information on the unemployment rate for the Ottawa-Gatineau region. The instructions for task 3 required participants to call United Airlines and find out if there was a flight from Toronto to New York (Kennedy airport) around 6:00pm tomorrow evening and write down the flight number and departure time of the flight. Similarly, task 4 asked participants to call Air Canada and gather information regarding flight availability from Toronto to Vancouver around 12:00 pm tomorrow. Participants were again asked to note down the flight number and departure time of the flight. For the Air Canada task, participants were made aware that at some point during their call, they would be asked to provide the three-letter code of the departure and arrival cities. The three-letter codes were provided to participants in the form of written instructions during their interaction with the system.

No time limits were imposed for the completion of the four tasks and redialling was only permitted if the phone line was busy or a wrong number was reached. Thus, recovery from errors was expected to happen within the same call. The task ended when participants indicated that they have obtained the information, hung up, or if a live operator came on line. The latter usually happened either because a large number of errors triggered an automatic referral to an operator or because the participants pressed the option to connect to an operator despite being told not to do so at the beginning of each task. On average, participants took about 20 min to complete all four tasks.

Levels in the level achieved variable were defined as the junction point in the IVR conversation where a response was required. The IVR tasks were administered in fixed order starting with Service Canada, which required going through 4 levels in order to complete the task. Statistics Canada, which required going through 5 levels for completion, United

Airlines required going through 12 levels and Air Canada required going through 13 levels. Participants' performance on the four IVR was scored according to the number of successfully completed IVR tasks (maximum 4), the sum of the level achieved on each IVR task (i.e., number of responses +1: 0-34), and the number of errors. Any response that brought participants to a different level of the system than the one required to complete the task was coded as an error, including responses that were not understood by the system, or the failure to provide a response when one was required (within the time allotted by the IVR system). The number of times instructions were repeated by each IVR system (repetitions) was also recorded. Following the completion of the four IVR tasks, participants were administered the WAIS-IV and the WMS-IV batteries.

Statistical Analyses

Multiple regression analyses were performed in order to examine the cognitive and memory abilities that best predicted participants' performance on the four IVR tasks. The significance level was set at .05. Logistic regression was performed on the number of tasks successfully completed because of the categorical nature of that variable; the variable consisted of five categories (0-4 tasks completed). However, we obtained almost identical results with linear regression so we present the linear regression results that allowed us to enter factors in hierarchical order and evaluate the unique contribution of age and cognitive measures. Sequential linear regression was performed on the remaining data to determine what variables best predicted the level achieved on IVR tasks, the number of errors as well as the number of times people repeated the same level or had to go back one level in a task.

Results

Descriptive statistics showed that 20.5% of participants were not able to complete any of the IVR tasks, 26.5% completed only one task, 31.9% completed two tasks, 17.8% completed three tasks, and only 3.2% were able to complete all four IVR tasks (see Table 1). A larger number of participants were able to complete Service Canada (52.4%) and Statistics Canada tasks (57.8%), compared to the United Airlines (21.6) and Air Canada tasks (24.9%).

We used analysis of variances to examine the differences in age education and FSIQ among participants who completed 0, 1, 2, 3 and 4 tasks. The Welsh test of equality of means was significant at the .05 level: $F(4, 39) = 2.99, p < .03$. Post hoc analyses (Tamhane, 1979) revealed that participants who were unable to complete any of the tasks were significantly older compared to participants who completed three of the tasks ($p < .01$). No other group differences emerged for age. The groups did not differ in level of education, $F(4, 180) = 1.16, p = .33$. The five groups were also different in terms of participants' FSIQ, $F(4, 180) = 10.77, p < .01$. Post hoc analyses (Scheffé) indicated that participants who did not complete any of the tasks had a significantly lower FSIQ compared to all other groups ($p < .01$ for all comparisons). No other differences among groups emerged.

Measures That Predict the Number of IVR Tasks Completed

Linear regressions were performed on the data to determine which cognitive measures best predicted the number of IVR tasks completed by participants. The age variable was skewed and reciprocal transformation was performed. Table 2 displays the correlations between the variables, the unstandardized regression coefficients (B) and

intercept, the standardized regression coefficients (β ; beta), and the semipartial correlations (sr_i^2), and R , R^2 , and adjusted R^2 after entry of all variables. Step 1, using only age in the equation, led to an adjusted $R^2 = .05$, $F(1, 183) = 10.24$, $p < .05$. Step 2, which added working memory, produced an adjusted $R^2 = .20$, $F(2, 182) = 24.67$, $p < .001$. The addition of working memory to the equation with age resulted in a significant increment in the adjusted R^2 . Step 3, which added auditory memory together with age and working memory improved the prediction of number of tasks completed, leading to an adjusted $R^2 = .28$, $F(3, 181) = 24.86$, $p < .001$. The addition of perceptual reasoning, verbal comprehension, processing speed and visual memory did not reliably improve the adjusted R^2 , $p = .16$.

Measures That Predict the Level Achieved Within the IVR Tasks

Sequential linear regression was performed on the data to determine which cognitive variables best predicted the level achieved on each IVR tasks (Table 3). Step 1, with only age in the equation, led to an adjusted $R^2 = .00$, $F(1, 183) = 2.53$, $p = .11$. Step 2, which added working memory in the equation, increased the adjusted $R^2 = .08$, $F(2, 182) = 9.22$, $p < .001$. The addition of auditory memory in step 3 improved the prediction of level achieved, adjusted $R^2 = .12$, $F(3, 181) = 9.27$, $p < .001$ and also resulted in significant increase in the adjusted R^2 . The addition of perceptual reasoning, verbal comprehension, processing speed and visual memory did not reliably improve the adjusted R^2 , $p = .27$.

Because the level achieved variable on IVR tasks used the total number of levels attained on all four tasks, this variable is weighted more heavily for the two airplane tasks compared to the two simpler information tasks. Separate analyses that used levels attained for task 1 and 2 and levels attained for task 3 and 4 led to similar conclusions. In these

analyses, age was no longer significant. Auditory memory was the best predictor of level achieved for task 1 and 2, and working memory was the best predictor for task 3 and 4.

Measures That Predict the Number of Errors Made During the IVR Tasks

Because there was a larger number of levels in tasks 3 and 4 and thus higher chance for committing errors during the interaction with these two systems compared to tasks 1 and 2, we completed separate analyses for these tasks. The conclusions were similar, thus we present the results for all four tasks combined. Sequential linear regression was performed on the data to determine which variables best predicted the number of errors made during the IVR interactions. The number of errors variable was transformed using a logarithmic transformation because it was significantly positively skewed. Table 4 displays the results of the analyses. Step 1, which added age in the equation led to an adjusted $R^2 = .05$, $F(1, 183) = 11.14$, $p = .001$. Step 2, which added working memory in the equation, led to an adjusted $R^2 = .07$, $F(2, 182) = 8.39$, $p < .001$. The addition of working memory to the equation with age resulted in a small but significant increment in the adjusted R^2 . Step 3, which added auditory memory in the equation together with age and working memory, improved the prediction of number of tasks completed with an adjusted $R^2 = .11$, $F(3, 181) = 8.83$, $p < .001$. The addition of perceptual reasoning, verbal comprehension, processing speed and visual memory did not improve the adjusted R^2 , $p = .22$.

Measures That Predict the Number of Times Instructions Had to be Repeated (Repetitions) and the Number of Times Participants Had to Return to Previous Level (Back Responses)

Sequential linear regressions was performed on the data to determine which variables best predicted the number of times instructions were repeated and the number of times participants returned to the previous level during the IVR interactions. These variables were

transformed using log transformation. R was not significantly different from zero at the end of all steps for both regressions. For the analyses of number of repetitions after step 4 with all IVs in the equation, adjusted $R^2 = .00$, $F(7, 177) = .761$, $p = .62$ indicates that none of the variability in the number of repetitions during IVR interactions was predicted by age, cognitive function and memory. For the analyses of number of back responses after step 4 with all variables entered in the equations, adjusted $R^2 = .01$, $F(7, 177) = .715$, $p = .66$ indicates that only 1 % of the variability was explained by the predictors.

Most Common Errors for Each Task

Some of the errors that led to the non-completion of IVR tasks were directly linked to the programming of the IVR applications. In the Statistics Canada task, participants were asked to obtain information on unemployment rates; however, the term labour force was substituted for unemployment rates for one of the levels without any explanation by the IVR program. This prevented about one quarter of participants from completing the task.

In the United Airlines IVR task, participants were instructed to gather information for a flight that leaves “tomorrow”. However, the system initially provided information for the available flights for “today”. Participants needed to remember what the initial instruction was and select the “tomorrow” option in the IVR menu. For the Air Canada IVR system, if the number “0” was not used in the airport code to replace the letter “Z” as instructed (instead of the intuitive “9” found on the keypad), participants found themselves in an IVR loop from which many could not recover, and thus needed to hang up.

When we listened to the audio-recordings of the participants’ interactions with the IVR systems, it was clear that participants were very motivated to succeed. They spent a long time trying to obtain the requested information. Nevertheless, 17 participants hung up while

interacting with the United Airlines IVR system and 62 participants abandoned the task while interacting with Air Canada.

Discussion

Older people report difficulties using IVR systems (Katz et al., 1997; Miller et al., 2011). The present study attempted to understand which cognitive changes occurring during aging could best explain these difficulties. We systematically examined the cognitive and memory factors that predict successful completion of real-life interactive voice response tasks using normed tests measuring cognitive and memory domains. A substantial number of participants were unable to complete any of the four IVR tasks (20.5%) and overall success rate for completing all four IVR tasks was very low (3.2%). Predictably, the tasks requiring fewer responses to complete (Service Canada and Statistics Canada with 52.4% and 57.8% completion rate, respectively) were much easier than the other two tasks requiring a longer sequence of responses (United Airlines and Air Canada with 21.6% and 24.9% completion rate, respectively). The results indicated that poorer auditory and working memory was associated with poor IVR performance. Our casual observation of the participants during their interactions with the IVR systems suggested that the design of IVR programs compounded these problems and led to significant frustration in older people.

Age accounted for 5% of the variance in the number of tasks completed by participants, the addition of working memory added 16% to the explained variance and auditory memory explained additional 8%. The entering of perceptual reasoning abilities, processing speed abilities, verbal comprehension abilities and visual memory into the equation did not significantly improve the model and explained only additional 3% of the variance. These results suggested that the best cognitive predictors of the number of tasks

completed by older adults in our sample were their working memory and auditory memory. We also evaluated participants' success on IVR tasks by using a measure of partial success, which we called level achieved. The same factors (age, working and auditory memory) explained approximately the same proportion of the variance.

Taken together, the results of our study indicated that adults over the age of 65 experience significant difficulties in interacting with IVR systems. These findings are in line with the results of a study, which examined the relationship between age and IVR performance, (Dulude, 2002), a result also observed by Sharit and colleagues (2003). Dulude's study compared older IVR users to younger users and found significant age effects. The oldest adults in our sample experienced significantly more difficulties completing IVR tasks and made more mistakes compared to younger participants. However working memory and auditory memory, independently predicted the number of tasks participants were able to complete, the level that they were able to achieve on IVR and the number of errors they committed during the interactions. Although increasing age is associated with a number of cognitive changes (Drag & Bieliauskas, 2010), subgroups of older people experience various rates of age-related cognitive decline and different cognitive domains are more susceptible to age-related decline. The design of the current study does not allow to draw a clear conclusion regarding the interplay of aging and cognitive decline on IVR performance.

Our findings that neither age nor cognitive and memory status were related to the number of repeat and back responses used by participants was somewhat surprising. Repeat and back menu options are programming strategies intended to help users recover from mistakes without having to call back the system. One possible explanation for our findings is that these recovery options were not always made available to users at each level of the

interaction. In many cases, the opportunity “to press star in order to repeat the menu or pound to go back” were mentioned at the first level of the IVR interaction but not at the higher levels. It is possible that participants forgot the previous instructions as they were focusing on the new information presented to them and were unable to use the options that may have potentially increased their success with the systems. This is supported by our data indicating that 26% of participants did not use the repetition options and 28% used it only once. The results for the back responses were similar in that 21% did not use the option and 20% used it only once. These results combined with the lack of success of many participants in the IVR tasks strongly suggest that participants would have benefited from being reminded of the error recovery options. Finally, there were clear indications that IVR programming could compound the difficulties in interacting with automated systems not only for older people but also for anyone with lower auditive or working memory abilities.

The interaction between the design shortcomings of the IVR tasks used in this experiment and the changes observed in cognitive abilities during aging presents a challenge to frame a theory that could relate the IVR experience as such to cognitive constructs and develop theories that could explain and predict who will have the most difficulties with IVR systems. A number of considerations are also relevant for the interpretation of the cognitive abilities necessary for IVR interactions. The first one, common to all telephone conversations, is the lack of visual cues that convey additional information and provide validation of a “right” or “wrong” answer. Secondly, since IVR systems use pre-recorded messages, they also remove communication elements embedded in voice intonation. This reduction in information density forces the IVR user to rely entirely on the semantic content of the pre-recorded messages. This reliance might explain the pre-eminence of auditory and

working memory in the present experiment and the relative lack of influence of the other cognitive domains. As eloquently discussed by Salthouse (2010), cognitive aging is not a simple overall decline in abilities but rather a multi-factorial process in which mediators and moderators may determine not only the rate of decline but also which cognitive domains decline (Salthouse, 2010).

Finally, a common occurrence in IVR interactions is frustration often manifested by the reversion to natural language (instead of the range of words requested by the IVR system). This frustration is possibly associated with the contrast between the usually intelligently sounding IVR voice and the IVR system often-moronic response to deviation of a predetermined script. On this level, the failure of people to interact with IVR systems may be directly related to the failure of IVR systems to interact with people.

Are IVR systems hopeless for older people? Most likely not. The main problem of IVR systems is that, usually, their programming does not allow for an easy recovery from errors – something that would be readily done in a real person interaction. However, error recovery is often intertwined with the specific parts of an IVR application and problems (some predictable, some less predictable) need to be resolved with the help of a representative sample of older people that the application is intended for. Secondly, IVR systems do not make any provision for people with different cognitive abilities; again this is something that would be done by a live operator. This lack of adaptability to users may be an important factor that leads to frustration and rejection of these systems. However, the hardware and software currently available does allow for the development of an adaptive approach to IVR. Adaptive approach to IVR could include the modulation of the volume and the speed of delivery of the automated interaction as well as alternate menus of shorter

length and fewer choices for older people. Finally, each IVR application presents with its unique challenges, and some of these challenges will be greater for older people with impaired auditory and working memory. In our current projects, we find that these challenges are not insurmountable. They can be overcome by programming IVR applications that adapt to people rather than asking people to adapt to IVR systems. Considering the potential use of IVR technology in health applications and the associated cost savings, this avenue needs to be explored further.

Funding: This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and by TelAsk Technologies.

Acknowledgements : We would like to thank Dwayne Schindler for statistical consultation and Halina Bruce for study coordination.

References

- Bardone, A. M., Krahn, D. D., Goodman, B. M., & Searles, J. S. (2000). Using interactive voice response technology and timeline follow-back methodology in studying binge eating and drinking behavior: Different answers to different forms of the same question? *Addictive Behaviors, 25*(1), 1-11. doi: 10.1016/S0306-4603(99)00031-3
- Corkrey, R., Parkinson, L., & Bates, L. (2005). Pressing the key pad: Trial of a novel approach to health promotion advice. *Preventive Medicine, 41*(2), 657-666. doi: 10.1016/j.ypmed.2004.12.008
- Corkrey, R., Parkinson, L., Bates, L., Green, S., & Htun, A. T. (2005). Pilot of a novel cervical screening intervention: Interactive voice response. *Australian and New Zealand Journal of Public Health, 29*(3), 261-264. doi: 10.1111/j.1467-842X.2005.tb00765.x
- Crawford, A. G., Sikirica, V., Goldfarb, N., Popiel, R. G., Patel, M., Wang, C., Nash, D. B. (2005). Interactive voice response reminder effects on preventive service utilization. *American Journal of Medical Quality, 20*(6), 329-336. doi: 10.1177/1062860605281176
- Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (create). *Psychology and Aging, 21*(2), 333-352. doi: 10.1037/0882-7974.21.2.333
- Czaja, S. J., Sharit, J., & Nair, S. N. (2008). Usability of the medicare health web site. [Letter]. *JAMA : The Journal of the American Medical Association, 300*(7), 790-792. doi: 10.1001/jama.300.7.790-b
- Drag, L. L., & Bieliauskas, L. A. (2010). Contemporary review 2009: Cognitive aging. [Review]. *Journal of Geriatric Psychiatry and Neurology, 23*(2), 75-93. doi: 10.1177/0891988709358590
- Dulude, L. (2002). Automated telephone answering systems and aging. *Behaviour & Information Technology, 21*(3), 171-184. doi: 10.1080/0144929021000013482
- Fiscella, K., Humiston, S., Hendren, S., Winters, P., Idris, A., Li, S. X., Marcus, S. (2011). A multimodal intervention to promote mammography and colorectal cancer screening in a safety-net practice. *Journal of the National Medical Association, 103*(8), 762-768. doi: PMid:22046855
- Forster, A. J., & van Walraven, C. (2007). Using an interactive voice response system to improve patient safety following hospital discharge. *Journal of Evaluation in Clinical Practice, 13*(3), 346-351. doi: 10.1111/j.1365-2753.2006.00702.x

- Granger, B. B., & Bosworth, H. B. (2011). Medication adherence: Emerging use of technology. *Current Opinion in Cardiology*, 26(4), 279-287. doi: 10.1097/HCO.0b013e328347c150
- Hasvold, P. E., & Wootton, R. (2011). Use of telephone and sms reminders to improve attendance at hospital appointments: A systematic review. *Journal of Telemedicine and Telecare*, 17(7), 358-364. doi: 10.1258/jtt.2011.110707
- Heisler, M., Halasyamani, L., Resnicow, K., Neaton, M., Shanahan, J., Brown, S., & Piette, J. D. (2007). "I am not alone": The feasibility and acceptability of interactive voice response-facilitated telephone peer support among older adults with heart failure. *Congestive Heart Failure*, 13(3), 149-157. doi: 10.1111/j.1527-5299.2007.06412.x
- Katz, M. J., Aspden, P., & Reich, W. A. (1997). Public attitudes toward voice-based electronic messaging technologies in the united states: A national survey of opinions about voice response units and telephone answering machines. *Behaviour & Information Technology*, 16(3), 125-144. doi: 10.1080/014492997119860
- Kim, H., Bracha, Y., & Tipnis, A. (2007). Automated depression screening in disadvantaged pregnant women in an urban obstetric clinic. *Archives of Women's Mental Health*, 10(4), 163-169. doi: 10.1007/s00737-007-0189-5
- Kim, H., Geppert, Joni, Quan, Tu, Bracha, Yiscah, Lupo, Virginia, & Cutts, Diana. (2012). Screening for postpartum depression among low-income mothers using an interactive voice response system. *Maternal and Child Health Journal*, 16(4), 921-928. doi: 10.1007/s10995-011-0817-6
- McPhail, G. L., Ednick, M. D., Fenchel, M. C., VanDyke, R., Chima, A., Amin, R. S., & Seid, M. (2010). Improving follow-up in hospitalised children. *Quality & Safety in Health Care*, 19(5), e35. doi: 10.1136/qshc.2009.034538
- Miller, D.I., Bruce, H., Gagnon, M., Talbot, V., & Messier, C. (2011). Improving older adults' experience with interactive voice response systems. *Telemedicine Journal and E-Health*, 17(6), 452-455. doi: 10.1089/tmj.2010.0204
- Moore, H. K., Mundt, J. C., Modell, J. G., Rodrigues, H. E., DeBrotta, D. J., Jefferson, J. J., & Greist, J. H. (2006). An examination of 26,168 hamilton depression rating scale scores administered via interactive voice response across 17 randomized clinical trials. *Journal of Clinical Psychopharmacology*, 26(3), 321-324. doi: 10.1097/01.jcp.0000219918.96434.4d
- Mundt, J. C., Bohn, M. J., King, M., & Hartley, M. T. (2002). Automating standard alcohol use assessment instruments via interactive voice response technology. *Alcoholism, Clinical and Experimental Research*, 26(2), 207-211. doi: 10.1111/j.1530-0277.2002.tb02525.x

- Mundt, J. C., Geralts, D. S., & Moore, H. K. (2006). Dial "t" for testing: Technological flexibility in neuropsychological assessment. *Telemedicine and e-Health, 12*(3), 317-323. doi: 10.1089/tmj.2006.12.317
- Park, H. L., O'Connell, J. E., & Thomson, R. G. (2003). A systematic review of cognitive decline in the general elderly population. *International Journal of Geriatric Psychiatry, 18*(12), 1121-1134. doi: 10.1002/gps.1023
- Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S., Mundt, J. C., Fava, M. (2006). An evaluation of the quick inventory of depressive symptomatology and the hamilton rating scale for depression: A sequenced treatment alternatives to relieve depression trial report. *Biological Psychiatry, 59*(6), 493-501. doi: 10.1016/j.biopsych.2005.08.022
- Salthouse, T. A. (2010). *Major issues in cognitive aging*. Oxford ; New York: Oxford University Press.
- Schroder, K. E., Johnson, C. J., & Wiebe, J. S. (2007). Interactive voice response technology applied to sexual behavior self-reports: A comparison of three methods. *AIDS and Behavior, 11*(2), 313-323. doi: 10.1007/s10461-006-9145-z
- Sharit, J., Czaja, S. J., Nair, S., & Lee, C. C. (2003). Effects of age, speech rate, and environmental support in using telephone voice menu systems. *Human Factors, 45*(2), 234-251. doi: 10.1518/hfes.45.2.234.27245
- Statistics Canada. (2009). Survey of household spending in 2008.
- Tamhane, A C. (1979). A comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association, 74*(366), 471-480. doi: 10.2307/2286358
- Treitz, F. H., Heyder, K., & Daum, I. (2007). Differential course of executive control changes during normal aging. *Aging, Neuropsychology, and Cognition, 14*(4), 370-393. doi: 10.1080/13825580600678442
- Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience and Biobehavioral Reviews, 26*(7), 849-857. doi: 10.1016/S0149-7634(02)00071-4
- Wechsler, D. (2008). *Wechsler adult intelligence scale, 4th edition*. San Antonio, TX: Pearson.
- Wechsler, D. (2009). *Wechsler memory scale, 4th edition*. San Antonio, TX: Pearson.

Table 1.

Participants' Age, Education, and Full Scale IQ as a Function of the Number of IVR Tasks They Were Able to Complete

Number of tasks completed	%	Age	Education	Full-scale IQ
0	20.5	$M=76.29, SD=7.54$ Range: 66-92	$M=13.26, SD=3.17$ Range: 7-18	$M=92.14, SD=10.84$ Range: 68-115
1	26.5	$M=73.29, SD=6.25$ Range: 65-87	$M=13.65, SD=2.46$ Range: 9-19	$M=102, SD=12.86$ Range: 68-123
2	31.9	$M=72.83, SD=6.26$ Range: 66-89	$M=14.03, SD=2.97$ Range: 9-22	$M=105.39, SD=13.55$ Range: 75-134
3	17.8	$M=71.06, SD=5.13$ Range: 65-83	$M=14.67, SD=3.15$ Range: 8-21	$M=107.12, SD=10.86$ Range: 88-128
4	3.2	$M=72.17, SD=2.6$ Range: 68-75	$M=14, SD=1.9$ Range: 12-16	$M=115.33, SD=9.85$ Range: 106-131

Table 2.

Sequential Regression of Age (Reciprocal), Cognitive Function and Memory on Number of IVR Tasks Completed

Variables	No. of IVR Tasks	Age (reciprocal)	WM	AM	PR	VC	PS	VM	B	SE B	β	R^2	Adjusted R^2	sr_1^2
Age	.23								46.74	70.85	.05	.05	.05	.05*
WM	.43	.17							.04	.01	.23	.21	.20	.16**
AM	.45	.31	.36						.02	.00	.30	.29	.23	.08**
PR	.36	.32	.40	.28					.01	.01	.15	.32	.29	.03
VC	.36	.16	.41	.41	.46				.00	.01	.07			
PS	.33	.45	.37	.35	.56	.40			.00	.01	.04			
VM	.31	.44	.31	.47	.59	.34	.46		-.00	.01	-.06			
Intercept	-2.85													
Means	1.57	.01	40.56	81.28	55.24	77.14	78.14	44.43						
SD	1.10	.001	6.81	20.00	15.18	17.02	17.84	14.90						
R^2	.32													
Adjusted R^2	.29													
R	.56*													

Notes. Because the age variable was entered as the reciprocal, the sign of the association between age and the other variables is positive rather than the expected negative. WM: Working memory; AM= auditory memory; PR= perceptual reasoning; VC= verbal comprehension; PS= processing speed; VM= verbal memory; B= unstandardized regression coefficient; SE B= intercept; β = standardized regression coefficient; sr_1^2 = semipartial correlation.

* $p < 0.05$ ** $p < 0.01$

Table 3.

Sequential Regression of Age (Reciprocal), Cognitive Function and Memory on Level Achieved on IVR Tasks

Variables	Level	Age	WM	AM	PR	VC	PS	VM	<i>B</i>	<i>SE B</i>	β	R^2	Adjusted R^2	sr_i^2
	Achieved on IVR	(reciprocal)												
Age	.12								-66.53	492.38	-.01	.01	.00	.01
WM	.30	.17							.14	.08	.14	.09	.08	.08**
AM	.31	.31	.36						.07	.03	.19	.13	.12	.04**
PR	.25	.32	.40	.28					.04	.05	.09	.16	.12	.03
VC	.30	.16	.41	.41	.46				.05	.03	.13			
PS	.23	.45	.37	.35	.56	.40			.01	.04	.04			
VM	.20	.44	.31	.47	.59	.34	.46		-.02	.05	-.04			
Intercept	4.47													
Means	21.03	.01	40.56	81.28	55.2	77.14	78.14	44.43						
SD	6.89	.001	6.81	20.00	15.1	17.02	17.84	14.90						
R^2	.16													
Adjusted R^2	.12													
R	.40*													

Notes. Because the age variable was entered as the reciprocal, the sign of the association between age and the other variables is positive rather than the expected negative. WM = working memory; AM = auditory memory; PR = perceptual reasoning; VC = verbal comprehension; PS = processing speed; VM = verbal memory; *B* = unstandardized regression coefficient; *SE B* = Intercept; β = standardized regression coefficient; sr_i^2 = semipartial correlation.

* $p < 0.05$ ** $p < 0.01$

Table 4.

Sequential Regression of Age (Reciprocal), Cognitive Function and Memory on Number of Errors

Variables	No. of Errors	Age (reciprocal)	WM	AM	PR	VC	PS	VM	<i>B</i>	<i>SE B</i>	β	<i>R</i> ²	Adjusted <i>R</i> ²	<i>sr</i> ²
Age	-.24								-30.87	17.66	-.14	.06	.05	.06**
WM	-.20	.17							-.00	.00	-.02	.08	.07	.03*
AM	-.31	.31	.36						-.00	.00	-.21	.13	.11	.04**
PR	-.25	.32	.40	.28					-.00	.00	-.12	.16	.12	.03
VC	-.28	.16	.41	.41	.46				-.00	.00	-.13			
PS	-.24	.45	.37	.35	.56	.40			.00	.00	-.02			
VM	-.20	.44	.31	.47	.59	.34	.46		.00	.00	.09			
Intercept	1.87													
Means	1.00	.01	40.56	81.28	55.24	77.14	78.14	44.43						
<i>SD</i>	.25	.001	6.81	20.00	15.18	17.02	17.84	14.90						
<i>R</i> ²	.16													
Adjusted <i>R</i> ²	.12													
<i>R</i>	.39*													

Notes. Because the age variable was entered as the reciprocal, the sign of the association between age and the other variables is positive rather than the expected negative. WM = working memory; AM = auditory memory; PR = perceptual reasoning; VC = verbal comprehension; PS = processing speed; VM = verbal memory; *B* = unstandardized regression coefficient; *SE B* = intercept; β = standardized regression coefficient; *sr*² = semipartial correlation.

* $p < 0.05$ ** $p < 0.01$

Study three entitled “Older people’s attitudes toward interactive voice response systems” evaluated seniors’ perceptions of the technology in relation to their success or failure to interact with IVR. The study also focused on sex differences in success in interacting with IVR and attitudes towards the technology. The study used the same sample of 185 older adults included in study two and provided further understanding regarding the use and usability of IVR in older adults. The study was accepted for publication in the journal of Telemedicine and E-Health (TMJ-2013-0028.R2). I was involved in the design of the experiment, recruitment of participants, testing of participants, analyzing the data and writing of the manuscript. France Aubé was involved in analyzing the data and writing and reviewing sections of the manuscript. Vincent Talbot from TelAsk Technologies reviewed versions of the manuscript and provided feedback. Dr. Michele Gagnon was involved in the role of a consulting neuropsychologist during the design stages of the project. She also provided training in neuropsychological test administration of the WAIS-IV and WMS-IV to the research assistants involved in the study. Dr. Claude Messier supervised the project and was the holder of the NSERC grant that funded the research.

Miller, D.I., Aubé, F., Talbot, V., Gagnon, M., & Messier, C. (2013). Older People’s Attitudes toward Interactive Voice Response Systems. *Telemedicine and E-Health (TMJ-2013-0028.R2)*.

Running head: FACTORS RELATED TO PERCEPTIONS OF IVR

Older People's Attitudes toward Interactive Voice Response Systems

Delyana Ivanova Miller, France Aubé, Vincent Talbot, Michèle Gagnon and

Claude Messier

Abstract

Interactive voice response systems (IVR) are computer programs, which interact with people to provide a number of services from business to health care. The healthcare applications are particularly relevant to older adults since they are the main consumers of most medical services. However, research has found that older adults experience significant difficulties with the IVR and have more negative attitudes towards the technology. Seniors' attitudes appeared to be related to their most recent experiences with IVR. The objective of this study was to examine attitudes towards four commercially available IVR and how these attitudes relate to participants' ability to interact with the technology in a sample of 185 (age 65+) community dwelling older adults. We also examined the effects of several demographic factors on both success and attitudes towards automated systems. We found a significant positive correlation between IVR success and attitudes towards IVR. However, a large subset of our sample gave higher ratings despite experiencing difficulties with the systems and these participants tended to have lower full-scale IQ. No gender differences emerged in terms of attitudes and ability to interact with IVR. Results also indicated that older adults in our sample viewed the IVR interaction as particularly taxing on their attention and concentration abilities.

Introduction

Interactive Voice Response (IVR) technology includes automated telephone answering systems, voice response units and automated attendants. In the last 20 years, there has been a rapid growth in the function of voice processing technology from simple voice mail, where messages can be stored and retrieved to complicated systems that direct incoming calls and provide access to a wide range of pre-recorded messages. The development of IVR technology has also created opportunities to enhance the quality of healthcare services. For example, IVR systems are currently used to provide reminders to patients for prescription refill, self-help treatments, reinforcement and support to patients, and information regarding chronic diseases management (Corkrey & Parkinson, 2002; Corkrey, Parkinson, Bates, Green, & Htun, 2005; Lee, 2003).

Older adults are generally slower at adopting new technology; however, in recent years the percentage of seniors using computers and the internet have significantly increased. According to Statistics Canada, the share of individuals aged 65 to 74 who use the internet had increased from 11% in 2000 to 28% in 2003. The same upward trend was evident among older adults aged 75 or older, but at a lower level (Statistics Canada, 2008). Despite the increasing percentage of older users, the age division in adoption of new communication technologies is still present (Czaja & Lee, 2007). When older adults are concerned, 98.9% of them own and use a telephone compared to only 33.1% who own a home computer, and 26.6% of them using internet at home (Statistics Canada, 2008). Thus, it appears that IVR systems, which use the telephone (a widely available and familiar device) as a medium of communication may have an advantage over other communication technologies in providing services to seniors.

The impact of aging on the adoption and use of new technologies seems to be related to a number of factors, such as age-related decrease in cognitive abilities, which makes the interaction difficult, anxiety in approaching technology, and negative attitudes towards new technology. Normal aging is associated with decline in a number of fluid intelligence abilities such as working memory, prospective memory, processing speed and slower acquisition of new skills and these have been shown to have important implications on performance when using unfamiliar technology (Czaja & Lee, 2007). In the case of IVR, the abilities that have been shown to impact performance in an older adult population are working memory and auditory memory (Miller, Gagnon, Talbot, & Messier, 2012).

In addition to having more difficulties in interacting with new technology compared to younger users, seniors have also been shown to have more negative attitudes towards new interfaces. In the case of IVR, two surveys of public attitudes towards these systems have reported that older users' perceptions of the technology are more negative compared to these of younger users and their attitudes seemed to be related to ones' most recent experience with IVR (Katz, Aspden, & Reich, 1997; Settle, Dillon, & Alreck, 1999). Further, a study examining older and younger women's performance on six IVR tasks and attitudes towards the technology also found a negative relationship between age and success in using the technology but the relationship between success with IVR and perceptions of the technology was not as strong as previously reported (older participants gave lower ratings to only two of the six systems used in the study) (Dulude, 2002). These results also indicated that older adults gave ratings comparable to younger adults to the other four systems despite having more difficulties in completing the IVR tasks. Similar results were reported in another case study evaluating ways to improve IVR design; two of the 14 subjects involved in the study

gave high usability ratings of the systems despite having experienced significant difficulties in their interaction (Schaffer & Sorflaten, 1996). Collectively, these results indicate that although, as we would expect, failure to accomplish one's goal when using technology is related to the perception of the technology, there is a subgroup of people for which this prediction does not hold. These findings may have important implications for IVR design that exploits users' feedback to improve interfaces. Most importantly, it would be helpful to identify subject characteristics that set apart these users from the rest.

In contrast to the popular belief that older adults are unwilling to use technology, research has found that seniors are interested in adopting new technology especially if they perceive it as helpful but they do experience more anxiety and are less confident in their ability to interact with it (Mrarquie, Jourdan-Boddaert, & Huet, 2002). Computer training have been shown to significantly improve older adults success in using new technology such as e-mail and the internet and increase their confidence in performing computerized tasks (Charness, Kelley, Bosman, & Mottram, 2001; Kubeck, Miller-Albrecht, & Murphy, 1999). It is also important to note that poor interface design and failure to account for the needs and preferences of users compound the problem and reduce the accessibility of new technology to seniors. Adopting a user centered approach in developing new applications has been proposed as a solution but very few applications are developed with users in mind and particularly older users (Czaja & Lee, 2007; Gould & Lewis, 1985).

There is also evidence for gender differences among seniors in the adoption of new communication technologies. More specifically, men appear to use technology more compared to women. In 2003, about one-third of men aged 65 to 74 reported using the internet or email compared with less than 25% of women. Further, men aged 75 or older

were more than twice as likely as women in the same age group to use computer technology (Statistics Canada, 2008). Some have related these differences to individual variability in spatial abilities, which appear to be important for successful navigation in computer mediated space (Terlecki & Newcombe, 2005) and these abilities have been shown to be higher in men compared to women (De Lisi & Cammarano, 1996; Masters, 1998). In the case of IVR technology, which use auditory instead of visual medium for interaction with users, the role of spatial abilities in optimal performance is not well established. One study has reported that special abilities predict successful interaction (Pak, Czaja, Sharit, Rogers, & Fisk, 2006), whereas another failed to find a consistent pattern between spatial abilities and IVR performance (Goldstein, Bretan, Sallnas, & Bjork, 1999). The gender differences in adopting computer technology may also be related to other factors such as gender stereotypes and anxiety related to use of technology. Women report less confidence and more anxiety regarding the use of computer software and both men and women regard this activity as masculine (Terlecki & Newcombe, 2005). In the case of IVR technology, the two surveys gathering data on IVR use and attitudes toward IVR have failed to find gender differences (Katz et al., 1997; Settle et al., 1999). Thus, it appears that the gender division found in the adoption and perceptions of other technological devices may not hold for IVR technology.

The goal of the present study was to evaluate participants' attitudes towards the four IVR systems included in our study. We looked at their overall attitudes towards the four IVR systems as well as participants' attitudes in relation to their success or failure to interact with the systems. Based of findings from previous research (Dulude, 2002; Katz et al., 1997), we hypothesized that participants' ratings of the systems would be influenced by their success in

using the IVR. However, there is evidence to suggest that some people are inclined to rate the systems high despite lack of success with IVR tasks (Schaffer & Sorflaten, 1996). Thus, we further examined our data for the presence of a subset of people who did not succeed with the tasks but reported positive attitudes towards IVR and examined the demographic variables that may have accounted for these differences. In addition, we evaluated our sample for sex differences in both success in using IVR as well as more positive or negative attitudes based on gender. Based on previous research (De Lisi & Cammarano, 1996; Masters, 1998; Terlecki & Newcombe, 2005), we hypothesized that men would be more successful in their IVR interaction and report more positive attitudes towards the four automated systems.

Method

Participants

The present study and all corresponding documentation was reviewed and approved by the Research Ethics Board of the University of Ottawa. One hundred and eighty-five (120 females: 65%) community dwelling people between 65 and 92 years of age (mean = 73.32 years, $SD=6.44$) were recruited from diverse socio-economic backgrounds, using advertisements in two free magazines for seniors and flyers, in community centers and subsidized housing buildings. Participants' education ranged from 7 to 22 years (mean = 13.88 years, $SD = 2.90$); 3% of participants had grade 8 or less, 14% had between grade 9 and grade 12, 23% had a high school diploma, 27% had some college or university, and 33% had a bachelor's, graduate, or professional degree. The only exclusion criteria were age younger than 65 and lack of proficiency in English. Participants were compensated \$100. The health questionnaire gathered self-reported demographic and health information. Eighty-

eighth percent were Caucasian. Ten percent reported being diabetic, 1.1 % reported having had a haemorrhagic stroke, 1.1% had been treated for a brain tumour, 1.6% reported another unspecified brain disease, and 0.5% had chronic hepatitis. Two percent reported currently seeing a psychiatrist, and 9.2% were currently being treated for depression. Thirty-five percent of the sample reported experiencing memory problems.

Measures

IVR tasks

The first two systems chosen were governmental IVR systems: Service Canada and Statistics Canada. Service Canada is the central agency that provides information on all government programs including old age pensions. Statistics Canada is the source of information regarding various aspects of Canada, many derived from census data. The last two were the IVR systems of United Airlines and Air Canada, which provide flight and ticket information to customers. The United Airlines system is a voice response system, while the other three systems require callers to use the telephone keypad to enter their responses.

Usability questionnaire The usability questionnaire adapted from Dulude's (2000) study is a qualitative measure of participants' experiences with the four IVR systems. The measure consisted of 10 short questions, which provided information regarding participants' perception of the usability of the four IVR systems (See Appendix A).

Wechsler Adult Intelligence Scale fourth edition

The Wechsler Adult Intelligence Scale fourth edition (WAIS-IV) comprises of 10 core sub-tests and 5 optional sub-tests, measuring a number of cognitive functions. The test provides four composite scores on the Verbal Comprehension scale, Perceptual Reasoning

scale, Working Memory scale and the Processing Speed scale as well as a global index of intellectual functioning-the Full Scale IQ. The WAIS-IV was normed on a population aged 16-90 years old (Wechsler, 2008). On average, it took 67 min to complete the 10 core subtests.

Wechsler Memory Scale fourth edition

The Wechsler Memory Scale fourth edition (WMS-IV) older adult battery (for people 65-90 years old) consists of five subtests, three with both immediate and delayed conditions. The battery produces four index scores: Auditory Memory Index, Visual Memory Index, Immediate Memory Index and Delayed Memory Index. The WMS-IV is normed on a sample of people aged 16-90 years (Wechsler, 2009). The battery was co-normed with WAIS-IV. The test took about an hour to complete.

Apparatus

Participants' interaction with the four automated systems over the telephone was recorded in order to facilitate scoring. A touch-tone phone (MITEL 5212) was used to call the IVR systems. The telephone used in this study had the keypad located on the base of the telephone and not on the headset. A Sony MP3 IC recorder (ICD-UX7 1F/UX81F) was attached to the phone line. Sony stereo headphones (MDR-XD200) were attached to the recorder. Thus, the examiners were able to listen to participants' interaction with the IVR as the interaction was unfolding.

Procedure

Upon arriving at the memory laboratory at the University of Ottawa, self-reported information on participants' health and memory status was obtained using a health

questionnaire. Next, participants completed four IVR tasks over the phone. The tasks asked participants to call an IVR system and obtain specific information. Participants were instructed to use only the automated menus and not use the option, which allowed them to speak to an operator. Instructions on the four different tasks were both verbally presented to participants and a list of these instructions was placed in front of them while they were completing the task. Participants were allowed to take notes while completing the tasks but most participants did not take advantage of this option.

The instructions for task one required participants to call Service Canada and obtain information on what one needs in order to qualify for an old age security pension. The instructions for task two asked participants to call Statistics Canada and obtain the latest information on the unemployment rate for the Ottawa-Gatineau region. The instructions for task three required participants to call United Airlines and find out if there was a flight from Toronto to New York (Kennedy airport) around 6:00pm tomorrow evening and write down the flight number and departure time of the flight. Similarly, task four asked participants to call Air Canada and gather information regarding flight availability from Toronto to Vancouver around 12:00 pm tomorrow. Participants were again asked to note down the flight number and departure time of the flight. For the Air Canada task, participants were made aware that at some point during the call, they would be asked to provide the three-letter code of the departure and arrival cities. The three-letter codes were provided to participants in the form of written instructions.

No time limits were imposed for the completion of the four tasks and re-dialling was only permitted if the phone line was busy or a wrong number was reached. Thus, recovery from errors was expected to happen within the same call. The task ended when participants

indicated that they have obtained the information, hung up, or if a live operator came on line. The latter usually happened either because a large number of errors triggered an automatic referral to an operator or because the participants pressed the option to connect to an operator despite being told not to do so at the beginning of each task. On average, participants took about 20 min to complete all four tasks.

The IVR tasks were administered in fixed order starting with Service Canada, which required going through four levels in order to complete the task. Statistics Canada, required going through five levels for completion, United Airlines required going through twelve levels and Air Canada required going through thirteen levels (See Appendix B). Following the completion of each IVR task participants were asked to fill out the usability questionnaire for that particular system. The usability questionnaires were identical for all four systems. Participants' performance on the four IVR was scored according to the number of successfully completed IVR tasks (maximum 4).

Following the completion of the four IVR tasks, participants were administered first the WAIS-IV and second the WMS-IV batteries in fixed order.

Analyses

The significance level was set at .05 and was adjusted for repeated tests. We used McNemar's tests to examine the differences in participants' performance on the four IVR tasks. A non-parametric test was chosen because of the categorical nature of the success variable (either completed the task or not). The alpha level was adjusted for the six comparisons. Alpha was set at .01. Independent samples t-test was used to examine differences between men and women's in success in using IVR and attitudes towards the four systems.

Results

Success in Completion of the Four IVR Tasks

Not all participants succeeded in completing the IVR tasks. Twenty-one percent of participants were not able to complete any of the tasks assigned and only 3.2% succeeded in completing all four IVR tasks. Seventeen percent of participants completed only one task, 32% completed two tasks and 17.8% completed three tasks. Statistics Canada was the IVR task with the highest rate of completion at 57.8% success rate, followed by Service Canada (52.4% success rate). The United Airlines system (the only voice-recognition system) had the lowest success rate at 21.6% and Air Canada had 24.9% success rate.

McNemar's test was used to compare participants' performance across the four different IVR tasks. There were no significant differences between participants' performance on the Service Canada task and the Statistics Canada task, and between the United Airlines task and the Air Canada task ($\chi^2(1, N = 185) = 1.066, p = .302$ and $\chi^2(1, N = 185) = .446, p = .504$, respectively). There were significant differences between participants' performance on the Service Canada task and the Air Canada task ($\chi^2(1, N = 185) = 27.473, p = .001$), between the Service Canada task and the United Airlines task ($\chi^2(1, N = 185) = 36.894, p = .001$), between the Statistics Canada and the United Airlines task ($\chi^2(1, N = 185) = 50.069, p = .001$), and between the Statistics Canada and the Air Canada task ($\chi^2(1, N = 185) = 41.379, p = .001$).

We used an independent samples t-test to examine sex differences in success with completing IVR tasks in our sample. There were no significant differences between the

performance of men ($M = 1.49$, $SD = 1.13$) and women ($M = 1.60$, $SD = 1.09$), $t(183) = .683$, $p = .50$.

Participants' Perceptions of the IVR Systems

We examined the data for sex differences in perceptions of the systems on overall ratings for the four IVR on the Usability Questionnaire using independent samples t-test. There were no significant differences between the ratings of men ($M = 130.51$, $SD = 19.82$) and women ($M = 127.13$, $SD = 20.33$), $t(183) = .109$, $p = .28$.

We also examined participants' perceptions of the four tasks in relation to their ability to complete each of the tasks by using their scores on the Usability Questionnaire and whether or not they were able to complete the tasks. We conducted independent samples t-tests. For the Service Canada IVR system, there was a significant difference in the ratings between the two groups: people who completed the task ($M = 37.76$, $SD = 6.84$) and those who were unable to complete the task ($M = 34.75$, $SD = 8.21$), $t(183) = 2.72$, $p < .01$. For the Statistics Canada IVR system, we found no significant differences in the ratings between the two groups: people who completed the task ($M = 35.67$, $SD = 7.71$) and those who were unable to complete the task ($M = 33.58$, $SD = 8.68$), $t(183) = 1.73$, $p = .09$, indicating that participants' success with the system did not influence their rating of the system.

There was a significant difference in the ratings between the two groups for the United Airlines task: people who completed the task ($M = 33.30$, $SD = 8.34$) and those who were unable to complete the task ($M = 28.58$, $SD = 7.55$), $t(183) = 3.42$, $p < .01$. Ratings were also significantly different for the Air Canada task: people who completed the task ($M = 30.09$, $SD = 7.07$) and those who were unable to complete the task ($M = 26.78$, $SD = 7.07$), $t(183) = 2.75$, $p < .01$.

System Usability Analyses

The Service Canada System had the highest and more favorable ratings by participants on the Usability Questionnaire ($M = 36.33$, $SD = 7.65$), followed by the Statistics Canada IVR system ($M = 34.79$, $SD = 8.17$), United Airlines ($M = 29.60$, $SD = 7.95$) and the Air Canada ($M = 27.60$, $SD = 7.19$). We conducted repeated measures ANOVA to evaluate the differences between means scored on the Usability Questionnaire for the four IVR systems. There was a significant between subjects effect $F(1, 184) = 68.870$, $p < .01$. The pairwise comparisons adjusted for multiple comparisons revealed that the ratings for all four systems were significantly different from each other ($p < .01$), except between the Service Canada and Statistics Canada IVR systems ($p = .02$) (see Table 1 for means and standard deviations).

Next, we examined the differences in ratings of the four systems for each of the 10 questions comprising the Usability Questionnaire using a non-parametric equivalent of related samples t-test, Wilcoxon Signed Rank Test. Alpha level was adjusted at the .01 level to account for multiple comparisons (0.05/6 comparisons). There were no significant differences in ratings for most questions between the Service Canada and Statistics Canada systems and between United Airlines and Air Canada. The only differences were between Service Canada and Statistics Canada against the two airline systems. On item five *The operator's voice was very clear*, only Service Canada was rated higher compared to all other three systems ($p < .01$). On item ten *I thought the operator spoke too quickly* there were no significant differences between the four systems.

Although, the IVR performance of participants was significantly correlated with the way they rated the systems (people who completed more tasks were likely to give higher

scores to the system, the correlation between IVR success variable and the IVR ratings variable was only .189, $p < .01$. Thus, we proceeded by dichotomizing the two variables in order to evaluate if our sample included people who did not succeed on the IVR tasks (completed 0,1 or 2 tasks) but rated the system high (overall score on the Usability Questionnaire higher than 120). The results revealed that half of the sample (50.8% of people) gave higher ratings to the systems despite having difficulties in completing the tasks. The rest of the sample included: 4.9% of participants who were successful with the systems but gave low ratings; 28.1% who had trouble with the IVR and also gave low ratings; and 16.2% who were successful and gave high ratings.

We ran a series of t-tests in order to evaluate the differences on demographic variables between the people who gave high ratings to the IVR despite their lack of success and the rest of our sample. The dependent variables included age, level of education, and self-reported memory status; none of these variables were significantly different between the two groups. The only difference emerged in participants' scores on intelligence testing; participants who gave high scores despite their lack of success tended to have lower FSIQ (composite score $M = 243.23$; $SD = 46.51$, vs. 260.37 , $SD = 39.66$), $t(183) = 2.693$, $p < .01$.

Discussion

Despite the increased popularity of IVR systems, research regarding the acceptability and usability of these systems especially in older adult population is limited. This study examined the impact of several demographic variables in a sample of community dwelling older adults aged 65 and older on both success in using IVR and their attitudes toward the technology.

Our hypothesis regarding gender differences in both success with IVR and attitudes regarding the technology was not supported. Men and women in our sample did not differ in the number of IVR tasks they were able to complete and our sample as a whole had significant difficulties interacting with the systems. Further, no gender differences emerged in participants' ratings of the IVR. Thus, our findings were in line with the two surveys which found no gender differences in attitudes towards automated systems (Katz et al., 1997; Settle et al., 1999). Taken together, our findings and the findings from the two surveys suggest that IVR technology is unique in respect to gender differences and how these relate to the way people view the technology and how successful they are in their interactions with it. We did not measure anxiety related to the use of IVR or familiarity with the technology -- two factors that have been previously reported to impact people's ability to interact with new technology (Terlecki & Newcombe, 2005); thus, we are unable to comment on their significance in the case of IVR.

The ratings of the four IVR systems followed the same pattern as the success rate of participants in their interactions with the systems. Not surprisingly, the two systems that were easier to navigate (Service Canada and Statistics Canada) were also rated higher compared to the two airline systems (United Airlines and Air Canada). These findings were also in line with previous research showing significant relationship between positive experience with IVR and perceptions of the technology (Dulude, 2002; Katz et al., 1997). Nevertheless, we noted that the strength of the correlation between success and overall ratings of the systems was somewhat low. Further evaluation of the data revealed that half of our sample gave much higher ratings of the systems than what would have been expected on the basis of their success with the IVRs; similar findings were also reported in previous

studies (Dulude, 2002; Schaffer, Sorflaten, Venkateswaran, & Miracle, 2000). In our sample, participants who gave higher ratings of the systems despite their general lack of success were only different in terms of their full-scale IQ from the rest of the sample. This finding has an important implication for IVR design that relies on customers' ratings in evaluating the usability of their systems. Designers should be mindful of the fact that a subpopulation of people with lower cognitive abilities may rate the systems high despite experiencing significant difficulties in using the technology. Our study design does not allow us to comment on other factors that may also be in play regarding these findings. However, some of the comments that participants made after their interactions with the IVRs led us to believe that many of them attributed their difficulties in using the technology to their personal abilities rather than a faulty IVR design.

Closer examination of the 10 questions comprising the Usability Questionnaire revealed that most items followed the same pattern as the one observed in terms of success rate with the four systems. No significant differences emerged between Service Canada and Statistics Canada and between United Airlines and Air Canada IVRs. However, ratings between the two easier systems and the two airline systems were significantly different. Participants gave much lower ratings of the two airline systems. It is important to note, however, that the one question concerned with the level of concentration required to complete the tasks, received low ratings (below the mean of 3) for all systems. This suggests that participants' experienced the IVR interaction (even for the two easier tasks) as taxing on their attention abilities. This finding is in line with our previous investigation which showed that working memory (a cognitive ability strongly related to attention) is a significant predictor of older people' ability to cope with IVR (Miller et al., 2012).

References

- Charness, N., Kelley, C. L., Bosman, E. A., & Mottram, M. (2001). Word processing training and retraining: effects of adult age, experience, and interface. *Psychological Aging*(16), 110-127.
- Corkrey, R., & Parkinson, L. (2002). Interactive voice response: Review of studies 1989-2000. *Behavior Research Methods, Instruments, & Computers*, 34(3), 342-353.
- Corkrey, R., Parkinson, L., Bates, L., Green, S., & Htun, A. T. (2005). Pilot of a novel cervical screening intervention: interactive voice response. *Aust N Z J Public Health*, 29(3), 261-264.
- Czaja, S. J., & Lee, C. C. (2007). The impact of aging on access to technology. *Universal Access in the Information Society*, 5(4), 341-349.
- De Lisi, R., & Cammarano, D. (1996). Computer experience and gender differences in undergraduate mental rotation performance. *Computers in Human Behavior*, 12, 351-361.
- Dulude, L. (2002). Automated telephone answering systems and aging. *Behaviour & Information Technology*, 21(3), 171-184.
- Goldstein, M., Bretan, I., Sallnas, E. L., & Bjork, H. (1999). Navigational abilities in aural voice-controlled dialogue structures. *Behaviour and Information Technology*, 18(2), 83-95.
- Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communication of the ACM*, 28, 300-311.
- Katz, M. J., Aspden, P., & Reich, W. A. (1997). Public attitudes toward voice-based electronic messaging technologies in the United States: A national survey of opinions about voice response units and telephone answering machines. *Behaviour & Information Technology*, 16(3), 125-144.
- Kubeck, J. E., Miller-Albrecht, S. A., & Murphy, M. D. (1999). Finding information in the World Wide Web: exploring older adults' exploration. *Education in Gerontology*, 25, 167-183.
- Lee, H., Friedman, M. E., Cukor, P., Ahem, D. (2003). Interactive voice response system (IVR) in health care services. *Nurs Outlook*, 51(6), 277-283.
- Masters, M. (1998). The gender difference on mental rotation test is not due to performance factors. *Memory and Cognition*, 26, 444-448.
- Miller, D., Gagnon, M., Talbot, V., & Messier, C. (2012). Predictors of successful communication with interactive voice response systems in older people. *The Journal*

of Gerontology, Series B: Psychological Sciences and Social Sciences,
doi:10.1093(available on-line).

- Mrarquie, J. C., Jourdan-Boddaert, L., & Huet, N. (2002). Do older adults underestimate their actual computer knowledge? *Behavior and Information Technology*, 21, 273-280.
- Pak, R., Czaja, S. J., Sharit, J., Rogers, W. A., & Fisk, A. D. (2006). The role of spatial abilities and age in performance in an auditory computer navigation task. *Computer and Human Behavior* 24(6), 3045-3051.
- Schaffer, E., & Sorflaten, J. (1996). Interactive voice response (IVR) case study. *Human Factors International*, available at: www.humanfactors.com.
- Schaffer, E., Sorflaten, J., Venkateswaran, M., & Miracle, G. (2000). Interactive Voice Response (IVR) Case Study. *The Journal of Electronic Commerce*, 12(2), 1-2.
- Settle, R. B., Dillon, T. W., & Alreck, P. L. (1999). Acceptance of phone-based interface for automated call direction. *Behaviour and Information Technology*, 18(2), 97-107.
- Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, 53(5-6), 433-441.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale, 4th Edition*. San Antonio, TX: Pearson.
- Wechsler, D. (2009). *Wechsler Memory Scale, 4th Edition*. San Antonio, TX: Pearson.

Table 1.

Usability Ratings for all Four IVR Systems (1-5).

Usability Questionnaire	Service Canada <i>M (SD)</i>	Statistics Canada <i>M (SD)</i>	United Airlines <i>M (SD)</i>	Air Canada <i>M (SD)</i>
<i>1. I found the system service easy to use</i>	3.90 (1.16)	3.79 (1.18)	2.89 (1.32)	2.65 (1.35)
<i>2. I felt very frustrated with this service*</i>	3.84 (1.19)	3.63 (1.27)	3.01 (1.34)	2.72 (1.35)
<i>3. I felt in control while using this service</i>	3.58 (1.24)	3.50 (1.19)	2.69 (1.26)	2.53 (1.26)
<i>4. I thought some of the choices are confusing*</i>	3.42 (1.22)	3.06 (1.35)	2.61 (1.26)	2.42 (1.27)
<i>5. The operator's voice was very clear</i>	4.57 (0.71)	4.30 (0.87)	4.23 (0.88)	4.29 (0.74)
<i>6. I am left with a good impression of that organization</i>	3.85 (1.09)	3.66 (1.14)	3.02 (1.32)	2.55 (1.18)
<i>7. I had to concentrate really hard in doing this task*</i>	2.85 (1.30)	2.77 (1.20)	2.37 (1.13)	2.24 (1.04)
<i>8. I sometimes felt lost while using this service*</i>	3.38 (1.34)	3.22 (1.31)	2.66 (1.28)	2.29 (1.17)
<i>9. I enjoyed using this service</i>	3.32 (1.78)	3.14 (1.16)	2.55 (1.25)	2.31 (1.18)
<i>10. I thought the operator spoke too quickly*</i>	3.63 (1.37)	3.72 (1.26)	3.58 (1.14)	3.61 (1.04)

Notes: The rating score for questions 2, 4, 7, 8 and 10 was inverted so that higher scores always indicate a more positive judgment. This was done so that the scores on all questions could be compared

Appendix A.

Usability Questionnaire
(adapted form Dulude, 2000)

On the scale from one to five shown bellow, indicate if you agree or disagree with the ten statements on this page.

1	2	3	4	5
strongly disagree	disagree	neutral	agree	strongly agree

Rating 1-5

1. I found the system service easy to use....._____
2. I felt very frustrated with this service....._____
3. I felt in control while using this service....._____
4. I thought some of the choices are confusing....._____
5. The operator's voice was very clear....._____
6. I am left with a good impression of that organization....._____
7. I had to concentrate really hard in doing this task....._____
8. I sometimes felt lost while using this service....._____
9. I enjoyed using this service....._____
10. I thought the operator spoke too quickly....._____

Complementary questions:

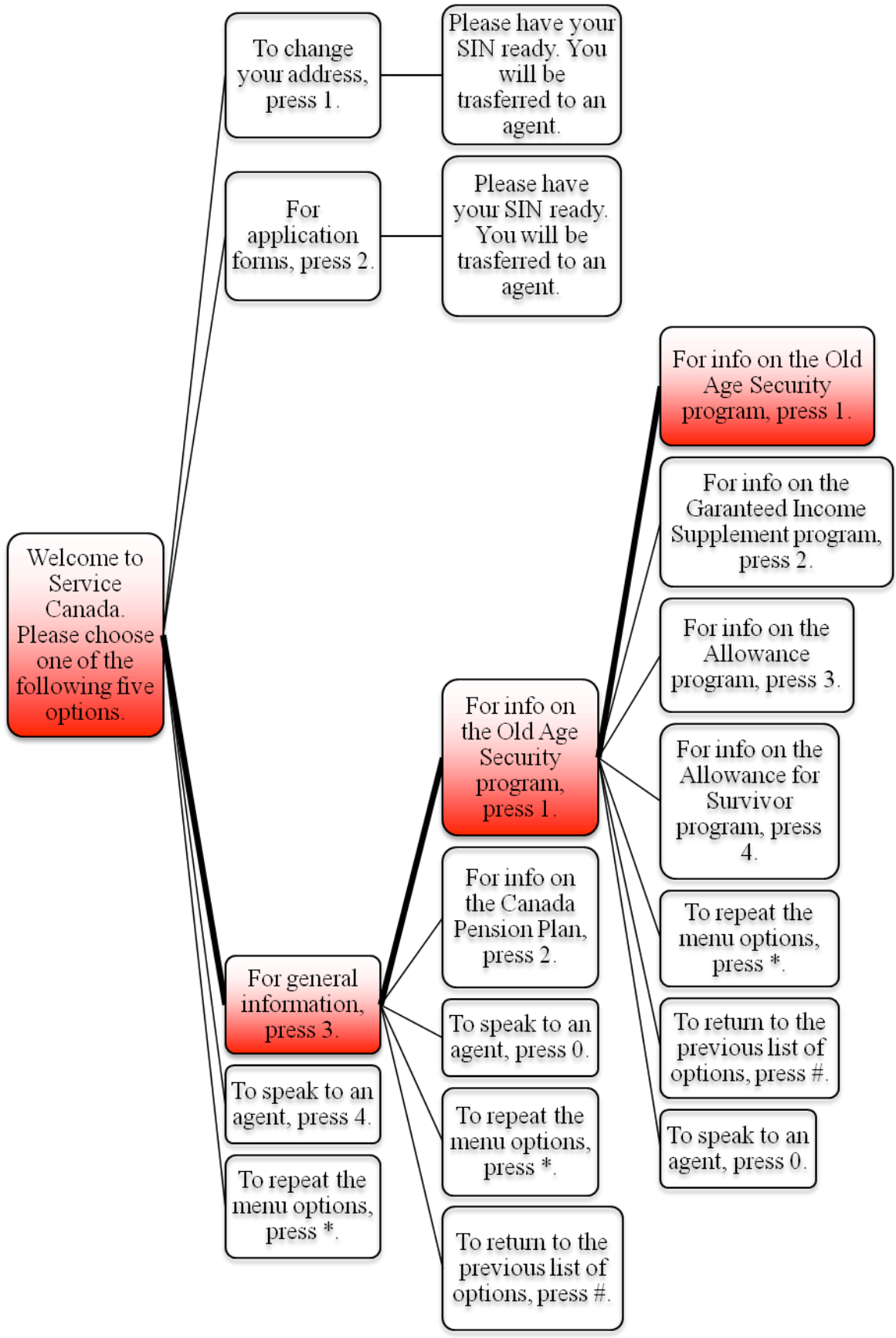
Had you ever used this system or a similar one before? How many times?

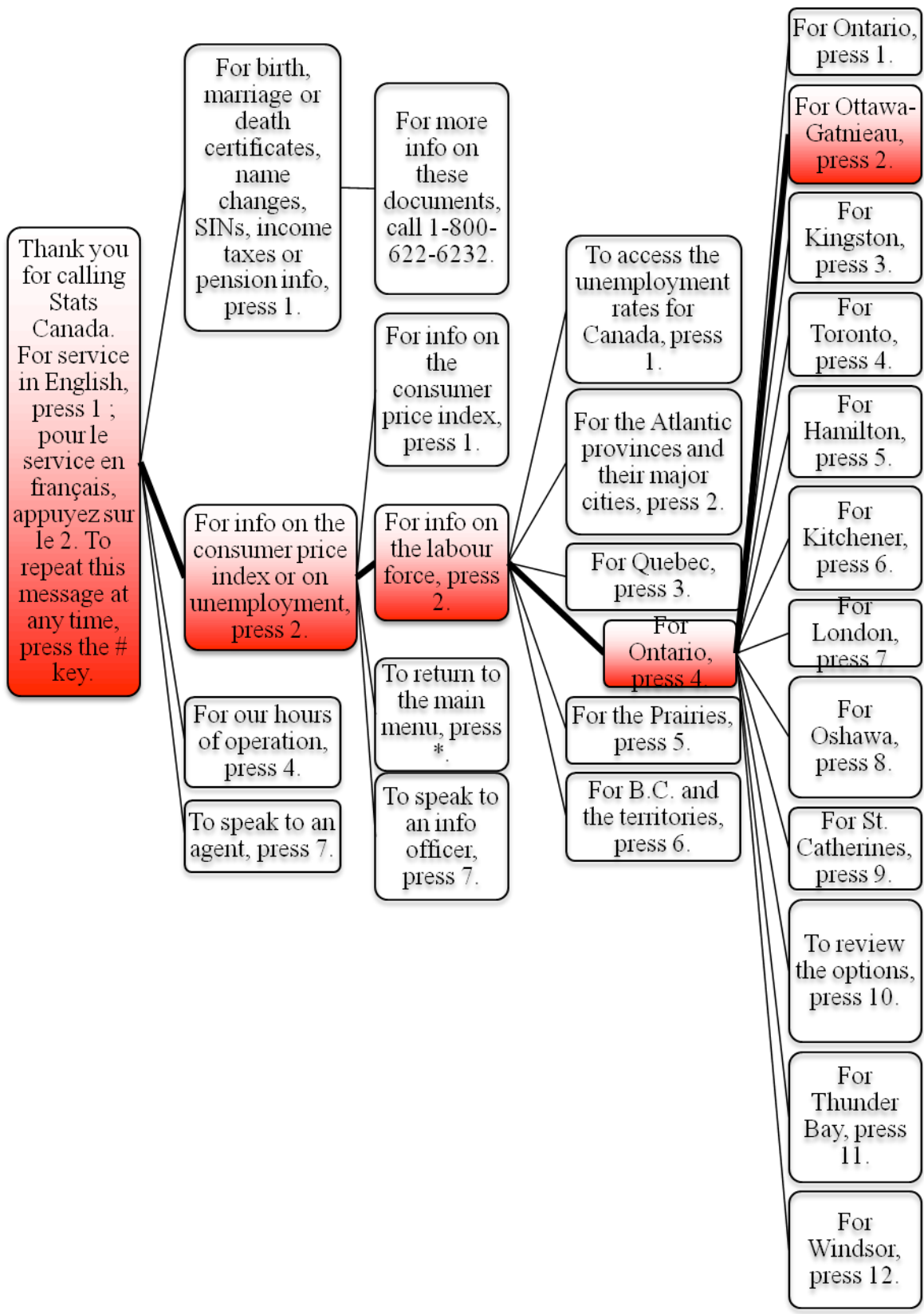
What was going on through your head when you were doing the task?

Do you have any other comments you would like to add about this task?

Appendix B

IVR Tasks





Thank you for calling United Airlines



Please select one of the following four options:



For departure and arrival information – Reservations – Mileage Plus – More options



Enter or say the flight number, or say “**I don’t know it**”.



Do you want information on arrival or **departure**?



You can interrupt me or say “start over” at any time.



What is the departure city? **Toronto**



What is the arrival city? **New York**



United Airlines service has more than one airport in New York. They are: New York, New

Jersey, **J.F.K., New York**, and La Guardia, New York.



The scheduled departure time? **6 p.m.**



Let’s make sure I got that right. (Provided information is repeated; however, the date is today’s

date). Is that right? Yes or **No**



Which one do you want to change? Departure city, arrival city, **date**, or time?



What date would you like? **Tomorrow**



I found two that you might want. Let’s find the right one.



Select flight number and departure time

Thank you for calling Air Canada



. For service in English **press 1**, pour continuer en français, faites le 2.



Please choose one of the following six options:

For flight arrival and departure information, press 1

To search for lowest fares, press 2 – For frequently asked questions, press 3

For reservations, press 4 – For assistance with Air Canada.com, press 5

For assistance with flight passes, press 6

To repeat this message, press 9.



If you know the flight number, press 1 – **If you don't know the flight number, press 2**



If you would like to hear the instructions on how to use the flight information system, press 1.

Otherwise, if you have a touch tone phone, press 2.

If at any time you need assistance or would like to start over, press *.



To hear arrival and departure information for today, press 1 – **For tomorrow, press 2**

For any other day, press 3.



Please enter the first 4 letters of the departure city or enter the 3 letter airport code followed by

the pound key. For Q or Z, press the 0 key: **TORO#** or **YYZ#**



For Toronto Island, press 1 – **For Toronto Pearson, press 2**



Please enter the first 4 letters of the arrival city or the 3 letter airport code followed by the pound

key. For Q or Z, press 0: **VANC#** or **YVR#**



Vancouver



To hear arrival times, press 1. – **To hear departure times, press 2.**



Enter the departure time followed by the pound key: **1200#**



For A.M., press 1 – For PM, press 2



Write down your flight information and departure time.

Study four entitled “Administration of a verbal fluency task and the digit span tasks using interactive voice response technology: validation and limitations” evaluated the feasibility in using voice-activated IVR for the purpose of administering and scoring neuropsychological tests. The study was a proof-of-principle examination and provided understanding regarding some of the challenges involved in using IVR technology for neuropsychological evaluations. The study employed a sub sample of 158 participants of the original sample of 185 older adults involved in study two and three. The study was submitted to *Frontiers in Teleneurology* and is currently under review. I was involved in the design of the experiment, recruitment of participants, testing of participants, analyzing the data and writing of the manuscript. Vincent Talbot from TelAsk Technologies was involved in the conceptualization, design and implementation of the IVR-COG. He also provided feedback on versions of the manuscript. Dr. Michele Gagnon was involved in the role of a consulting neuropsychologist during the design stages of the project. She also provided training in neuropsychological test administration of the WAIS-IV and WMS-IV to the research assistants involved in the study. Dr. Claude Messier supervised the project and was the holder of the NSERC grant that funded the research.

Running head: IVR ADMINISTRATION OF VERBAL FLUENCY AND DIGIT SPAN

Administration of a verbal fluency task and the digit span tasks using interactive voice
response technology: validation and limitations

Delyana Miller, Vincent Talbot, Michèle Gagnon and Claude Messier

Abstract

Interactive voice response systems (IVR) are computer programs, which interact with people using the telephone as a medium to provide a number of services from business to healthcare. We examined the ability of an IVR system to administer and score a verbal fluency task (fruits) and the digit span forward and backward in a sample of 158 community dwelling people aged between 65 and 92 years of age (full-scale IQ ranging from 68 to 134). Only 1% of participants were unable to complete all tasks mostly due to early technical problems in the study. Participants were also administered the Wechsler Adult Intelligence Scale fourth edition (WAIS-IV) and the Wechsler Memory Scale fourth edition (WMS-IV). The IVR system correctly recognized 90% of the fruits in the verbal fluency task and 93-95% of the number of sequences in the digit span. By and large, the IVR system underestimated the performance of participants because of recognition errors. In the digit span, these errors led to erroneous discontinuation of the test; however, the correlation between IVR scoring and clinical scoring was still high (correlations ranged between .93 and .95). The correlation between the IVR verbal fluency and the WAIS-IV Similarities sub-test was .31. The correlation between the IVR digit span forward and backward and the in-person administration was .46. We discuss how valid and useful IVR systems are as means of administration and scoring of neuropsychological tests.

Introduction

Interactive voice response systems (IVR) are generally described as computer systems interacting with people who, in turn, use the telephone keypad or speech to give answers to computer prompts. The use of IVR in psychological and neuropsychological assessment is a relatively new development.

Cognitive testing over the telephone has been used for over a decade. In the early applications, a “real” person performed both the administration and the scoring of the tests. These include the Telephone Adaptation of the Mini Mental State Examination (T3MS) (Norton et al., 1999), the Telephone Screening of Cognitive Status (TICS) (Brandt, Spencer, & Folstein, 1988), and its modified version (TICS-m) (Welsh, Breitner, & Magruder-Habib, 1993), used for dementia screening purposes. Other examples include the Minnesota Cognitive Acuity Screen (MCAS) for screening of dementia (Knopman, Knudson, Yoes, & Weiss, 2000), the Telephone-Administered Cognitive Test battery (TACT) (Prince et al., 1999) and the Indiana university telephone-based assessment of neuropsychological status (Unverzagt et al., 2007). Various studies that examined the validity of these telephone-administered tests reported good validity (Crooks, Petitti, Robins, & Buckwalter, 2006; de Jager, Budge, & Clarke, 2003; Knopman et al., 2010; Plassman, Newman, Welsh, Helms, & Breitner, 1994). The use of such tests for dementia screening may be appropriate in population studies but may not be as useful to establish more precise diagnosis such as mild cognitive impairment or sufficient to make a final diagnosis of dementia (Knopman et al., 2010). Researchers have expressed concerns about the construct validity of the adapted tests and suggested that tests administered by a technology platform (WEB, computer, telephone) may measure additional and/or different cognitive abilities (Bauer et al., 2012).

An example of an assisted computerized assessment is the Computer Assisted Telephone Interview (CATI) (Buckwalter, Crooks, & Petitti, 2002), which adapted the TICS-m for a computerized telephone administration. In a few instances, the computer program performs the administration and the scoring of the tests. An example of such system is the Computer Automated Dementia Screening instrument (IVR-ADS) (Mundt, Ferber, Rizzo, & Greist, 2001).

In the present experiment, we report on the development and preliminary validation of IVR algorithms, which administer and score in real time three commonly used neuropsychological tests using voice as the sole interaction mode. The goal of the present study was to evaluate the feasibility of the IVR application in older adults. In the course of this study, participants were also administered the WAIS-IV and WMS-IV in order to obtain an estimate of the validity of the IVR versions of the verbal fluency task (naming fruits), and the Digit Span Forward and Backward.

Method

Participants

The present study and all corresponding documentation was reviewed and approved by the Research Ethics Board of the University of Ottawa. One hundred and fifty-eight (105 females: 66.5%) community dwelling people between 65 and 92 years of age (mean = 72.97 years, $SD = 6.23$) were recruited from diverse socio-economic backgrounds, using advertisements in two free magazines for seniors and flyers, in community centers and subsidized housing buildings. Participants' education ranged from 7 to 21 years (mean = 13.85 years, $SD = 2.79$); 2% of participants had grade 8 or less, 14% had between grade 9 and grade 11, 34% had a high school diploma, 18% had some college or university, and 32%

had a bachelor's, graduate, or professional degree. Full-scale IQ ranged from 68 to 134. The only exclusion criteria were age younger than 65 and lack of proficiency in English.

Participants were compensated \$100. Self-reported demographic and health information was gathered using a health questionnaire. Eighty-eighth percent were Caucasian, 9.5% reported being diabetic, 0.6 % reported having had a haemorrhagic stroke, 1.3% had been treated for a brain tumour, 1.3% reported another unspecified brain disease, and 0.6% had chronic hepatitis, 2.5% reported currently seeing a psychiatrist, and 10.8% were currently being treated for depression. Sixty percent of the sample reported experiencing memory problems.

Measures

We used the Wechsler Intelligence Scale third edition (WAIS-III) version of Digit Span (Wechsler, 1997) for IVR administration of the Digit Span tasks and the WAIS-IV version (Wechsler, 2008) for the in-person administration in order to reduce practice effects. In the WAIS-IV, the Digit Span tasks were revised to include an additional section to the subtest, the Sequencing Digit Span. Some of the numbers in the Digit Span Forward and Backward were changed in order to eliminate similar sounding stimuli. The task has been found to be sensitive to age-related cognitive decline and a number of neurological disorders (Lezak, Howieson, & Loring, 2004).

The Animal Naming Task is a semantic fluency task. The task asks the examinee to name as many animals as the person can think of in 1 min and measures person's ability to rapidly generate words in response to a semantic cue (animals) (Lezak et al., 2004). Several alternate forms of the task have been developed and normed: naming vegetables, fruits, foods and clothing (Acevedo et al., 2000; Lucas et al., 1998; Spreen & Strauss, 1998). Semantic

fluency has been shown to significantly decline with age (Brucki & Rocha, 2004; Lucas et al., 1998).

The WAIS-IV 10 core subtests (Block Design, Similarities, Digit Span, Matrix Reasoning, Vocabulary, Arithmetic, Symbol Search, Visual Puzzles, Information and Coding) were administered by trained psychometrists following the standard clinical administration described in the test documentation. The WMS-IV, older adults version consists of 7 subtests: Logical Memory immediate and delayed recall, Visual Reproduction immediate and delayed recall, Verbal Pairs immediate and delayed recall and the Symbol Span.

The word recognition engine used for the development of our IVR was the Nuance Open Speech Recognizer version 3.0.3. There was no prior training of the system to optimize speech recognition of individual participants. A female professional voice talent recorded all the instructions and statements produced by the computer. The algorithms for the IVR tasks were developed by TelAsk technologies and two versions of the automated system were piloted using focus groups (Miller, Bruce, Gagnon, Talbot, & Messier, 2011). The subsequent versions of the IVR tasks were improved using the feedback from participants taking part in the focus groups. The final versions were used in this study. In addition to the three neuropsychological tests, the system allowed for the option of adjusting the volume and the speed of the conversation at the beginning of the interaction as well as for repetition of all instructions. There were five levels for the volume and the default was set at the medium volume 3. The speed adjustment had three levels and the default was set at 2.35 words/sec. The two other levels were slower (approximately 1.78 and 1.6 words/sec). These two slower levels were not digitally modified: rather the voice talent changed her speech rate in these recordings. Participants' interactions with the IVR were audio recorded.

The three tasks were scored by the IVR system and by a clinician (using the recordings) in order to examine the reliability of the system's scoring.

Apparatus

A touch-tone phone (MITEL 5212) was used to call the IVR systems. A Sony MP3 IC recorder (ICD-UX7 1F/UX81F) was attached to the phone line and Sony stereo headphones (MDR-XD200) was attached to the recorder. Thus, the headphones allowed the examiners to listen to participants' interaction with the IVR system as the communication was unfolding.

Procedure

The verbal fluency task was presented first; the examinees were instructed to name as many fruits as they could think of. They were also told that they have 1 min to complete the task. The Digit Span Forward task was presented second and the Digit Span Backward was presented last. For the two Digit Span tasks, we used the standard administration instructions from the WAIS-III manual. We also introduced a beep in the IVR version, which served as a cue to indicate that the string of numbers had been presented and participants needed to generate a response. This cue was introduced after pilot administrations revealed that many participants were waiting for additional digits and hesitant to start responding. We realized that the lack of a verbal (or non-verbal) cue for the last number, which would have been present in the in-person administration, led to uncertainty as to when the string of numbers had ended on each trial.

Upon arriving at the memory laboratory at the University of Ottawa, self-reported information on participants' health and memory status was obtained using a health questionnaire. Next, participants completed the IVR tasks over the phone. Participants were

not allowed to take notes during the interaction. In addition, all participants were administered the WAIS-IV and WMS-IV (older adult) batteries.

Results

Out of the 158 participants involved in the study, 152 completed all three tasks, one participant failed to complete the Digit Span Forward task and 6 participants completed only the verbal fluency task. The seven participants who completed only one or two tasks had difficulty interacting with the system, which caused the system to discontinue the conversation. Eleven percent of participants used the adjustment of the volume option; with 6 percent of them using the option to increase the volume and 5 percent of them used it to decrease the volume of the conversation. Fifteen percent of participants used the slowing down the speed of the conversation option. Eleven percent of participants asked the system to repeat the instructions for the verbal fluency task and 9 percent of participants did so for the digit span forward task.

The procedure and data analysis of this experiment were designed to answer three main questions. The first one was to determine if an IVR system could independently administer some simple neuropsychological tests and what were the problems associated with this administration. The second question was to determine if an IVR system could score the responses provided by the participants with accuracy and, if not, what were the technical hurdles that reduced accuracy. The final question, the most interesting to professionals that administer these tests, was whether neuropsychological tests administered over the phone using an IVR system provide comparable results to in-person administration with the caveat that for two of these tests (digit span forward and backward), the in-person administration followed the IVR administration (there was no counter-balancing of the order). The

following is a description of each test and a detailed analysis of the achievements and limits of each IVR test.

Verbal Fluency Task

Administration of the task

All participants completed the verbal IVR fluency task and no problems occurred during the administration. The fluency task involved little interaction between the IVR system and the participant beyond adjusting the speed of delivery and sound level or requesting repetition of instructions.

Scoring of the task

The correlation between the IVR verbal fluency task scored by the IVR system and scored by a clinician was $.89, p < .01$. The mean number of words scored by a clinician was higher ($M = 10.68, SD = 4.28$) compared to the scoring by the system ($M = 9.68, SD = 3.66$). The system was able to correctly recognize 90% of the fruit that participants named (in total participants named 1716 words, and the system identified 1539 words). Participants named 45 exotic fruit that were not part of the initial list entered into the system (e.g. breadfruit, cactus fruit, paw paws fruit, dragon fruit etc). The system failed to recognize 132 fruits (this number excludes the fruit that were not on the list: an error rate of 8%). Qualitative examination of the data revealed that the system's failure to correctly recognize some of the words was due to language issues such as non-English accent, poor pronunciation and verbal behaviours such as coughing, speaking inaudibly (very quietly), saying "umm", mumbling and speaking in full sentences to the system.

The clinician and IVR scoring of the test were identical in only 41.1% of cases (see Table 1). In 5.7% of cases, the system gave one point higher score indicating that some of

the repetitions or non-fruit words were identified as a correct response by the system. In the majority of cases (53.2%), the system gave a lower score than the one assigned by a clinician. For 27% (N=25) of people who had discrepancy in scores between clinician and system scoring the difference was due to fruits that were not part of the list that was entered into the system. Eleven percent (N=10) of the people with discrepant scores had foreign accents and 10% (N=9) had pronunciation difficulties. For the rest of participants (5.2%), the discrepancy was likely due to verbal behaviours described earlier.

Comparison with in person administration

Unfortunately, at the design phase of the study, we did not include a variation of the verbal fluency task (e.g. animals). Thus, we compared participants' performance on the IVR Verbal Fluency task to their performance on the Similarities subtest of the WAIS-IV administered in person. The Similarities subtest is the closest subtest in terms of cognitive functions being assessed (abstract verbal abilities). For all participants who completed the IVR Verbal Fluency task, the correlation between their IVR generated score and the Similarities subtest was $r = .32, p < .01$. The correlation between the clinician scoring of the Verbal Fluency task and the Similarities subtest was $r = .31, p < .01$. When we excluded the participants for whom the system and clinician scoring were discrepant, the correlations between the IVR Verbal Fluency task and the WAIS verbal subtest were the Similarities subtest was $r = .47, p < .01$.

Digit Span Forward

Administration of the task

For both the Digit Span forward and backward, participants had no problems understanding and following the IVR instructions of the task. The administration of the IVR

tests depended on the accurate recognition by the IVR system of each response to determine whether to continue with a number presentation in the same level or a number presentation for the next higher level. In the case where the IVR system did not recognize both sets in one level or when participants missed the first set of numbers and the IVR system did not recognize the second set, the test was incorrectly terminated.

Scoring of the task

In the following discussion, scores refer to the number of digit sequences correctly repeated. The correlation between the Digit Span forward task (adapted from the WAIS-III) scored by the IVR system and scored by a clinician was $r = .95, p < .01$. The mean number of digit sequences correctly repeated was ($M = 8.80, SD = 3.10$) for the clinician scoring and ($M = 8.34, SD = 3.13$) for the system scoring. The system was able to correctly recognize 95% of the digit sequences repeated by participants (participants named 1337 sequences and the system correctly identified 1268 of them). The system and clinician scoring of the IVR Digit Span forward task was identical for 63.8% of the cases (see Table 2). In 3.3% of cases, the system gave a higher score than the one assigned by the clinician. In the rest of the cases (32.9%), the system gave lower score than the one assigned by the clinician. The most common causes of failure on the part of the IVR to recognize a response are presented in Table 3. In many instances, the system by virtue of recognising one out of two strings of number of the same length allowed the participant to continue to the next level. However, the system discontinued the task for 12.5% ($N=19$) of participants earlier than it should have. One third of participants for whom the task was discontinued early ($N=7$) had foreign accent or difficulties with pronunciation. In addition, a number of people had a combination of

issues while completing the task that may have caused verbal recognition difficulties and led to discrepancy in scoring between clinician and the IVR system (see Table 4).

Comparison with in-person administration

For all participants who completed the IVR Digit Span forward, the correlation between their IVR scores and the in-person administration of the task was $r = .46, p < .01$. The correlation between clinician scored IVR Digit Span forward and the in-person administration of the task was $r = .48, p < .01$. For the 63.8% of participants for whom the IVR and clinician scoring of the Digit Span forward task was identical, the correlation between their IVR Digit Span forward (adapted from the WAIS-III) and the in-person administered Digit Span forward (WAIS-IV) was $r = .41, p < .01$. The mean number of digit sequences repeated correctly in the in-person administration of the Digit Span forward was $M = 9.58, SD = 2.30$, which was significantly higher than the IVR Digit Span forward score ($M = 8.80, SD = 3.10$) as indicated by paired samples t-test ($t(96) = 3.99, p < .01$). One contributor of this difference was the early termination of the task for some participants.

Digit Span Backward

Scoring of the task

The correlation between the Digit Span backward (adapted from the WAIS-III) scored by a clinician and by the IVR system was $r = .94, p < .01$. Once again, the mean number sequences repeated backwards was higher when scored by a clinician ($M = 6.07, SD = 2.33$) compared to the system scoring ($M = 5.63, SD = 2.18$). The system was able to recognize 93% of the digit sequences repeated backward by participants (participants repeated 928 sequences and the system recognized 861 of them). The system and clinician scoring of the IVR Digit Span backward task was identical in 68% of cases (see Table 5). There was only

one participant to whom the system assigned a higher score than the one assigned by the clinician. For the rest of participants (31.3%) the system gave a lower score. Examination of the audio recordings revealed the same issues as the one described for the Digit Span forward task (see Table 6). In addition, for 35% of participants who had discrepancy in scoring between a clinician and the system, the presentation of the Digit Span backward task was discontinued early. Only two participants for whom the task was discontinued early had accent and pronunciation difficulties. A number of people had a combination of issues while interacting with the system that likely led to recognition problems (see Table 7).

Comparison with in-person administration

For all participants who completed the task, the correlation between the IVR generated scores for the Digit Span backward and the in-person administration of the task was $r = .46, p < .01$. The correlation between clinician generated scores and the in-person administration of the task was $r = .50, p < .01$. For the 68% of participants for whom the IVR and clinician scoring of the Digit Span backward task was identical, the correlation between their IVR Digit Span backward (adapted from the WAIS-III) and the in-person administered Digit Span backward (WAIS-IV) was $r = .52, p < .01$. The mean number of digit sequences repeated correctly in the in-person administration of the Digit Span backward was ($M = 8.59, SD = 2.19$), which was once again significantly higher than that mean number of digit sequences obtained in the IVR version ($M = 6.07, SD = 2.33$) as indicated by paired samples t-test ($t(103) = 13.98, p < .01$).

Results for the Three IVR Tasks Combined

We also examined how many people had discrepancy in scoring on one or more of the IVR tasks. Frequency analyses revealed that only 20.9% of participants had identical

clinician and IVR generated scores for all three tasks, 39.9% of participants had discrepancies in scoring for one of the tasks, 30.7% had differences in scoring for two of the tasks and 8.5% of people had discrepant scores on all three tasks. We conducted one-way ANOVAs in order to examine if the different groups (people who had discrepancies on none, one, two, or all three tasks) were different in terms of their age, level of education, verbal and perceptual abilities, full-scale IQ, verbal and visual memory, and immediate and delayed memory. The ANOVAs were not significant for any of these variables.

In order to determine the cognitive functions that the three IVR tasks measured, the scores obtained on the IVR tasks were correlated with the WAIS-IV subtests results but we included only the participants for which the IVR system and the clinician scores were identical (see Table 8).

The IVR verbal fluency task was significantly correlated with all WAIS-IV verbal tasks: Similarities, Vocabulary and Information. The Digit Span forward was significantly correlated with the WAIS-IV Digit Span forward, Digit Span backward and the Information subtests. The Digit Span backward correlated with a number of tasks from the WAIS-IV measuring not only working memory but also verbal abilities and processing speed.

In order to compare how well the IVR administered Digit Span forward and backward tasks compare to the in-person administration of the tasks we also generated a correlation matrix of participants' scores on the ten core subtests of the WAIS-IV (see Table 8). The most striking differences were the higher correlation between the Digit Span forward task and the Digit Span backward task for the in-person administration compared to the correlation of the IVR administration (InP: $r = .42, p < .01$; IVR: $r = .20, p < .05$), higher correlation of Digit Span forward with the Arithmetic task for the in-person administration

(InP: $r = .29, p < .01$; IVR: $r = .03, ns$), and higher correlation for the in-person administration of Digit Span forward and the Vocabulary subtest (InP: $r = .26, p < .01$; IVR: $r = .18, ns$). The correlation between the in-person administration of the Digit Span backward task were higher with the Digit Span forward (InP: $r = .42, p < .01$; IVR: $r = .32, p < .01$) and higher correlation with the Matrix subtest for the in-person administration of the Digit Span backward (InP: $r = .23, p < .01$; IVR: $r = .07, ns$). In secondary analyses (data not shown), we examined whether the participants who slowed down the speed of the IVR conversation, adjusted the volume or repeated the instructions for the tasks had lower overall scores on the in-person administered cognitive batteries. No significant differences were found.

Principal component analyses with a varimax rotation was used on the data, and the variables included the three IVR administered tests (Verbal fluency, Digit Span Forward and Digit Span Backward) and the raw scores of the 10 core subtests of the WAIS-IV. Eigenvalues indicated that the first three factors explained 32.9, 13.7 and 9.7 percent of the variance respectively, for a total of 56% of the variance, which was also the case after varimax rotation was performed on the data. Thus, a three-factor solution was retained (see Table 9). The IVR Verbal Fluency task loaded on the second factor .57, together with all verbal tasks of the WAIS-IV (Similarities - .72, Vocabulary - .77 and Information - .78) providing support for the strong verbal component of the task. The IVR Digit Span forward and backward loaded on the third factor (.70 and .69, respectively) together with the Digit Span forward and backward from the WAIS-IV (.74 and .75, respectively). It was not surprising that the forward and backward versions of the task clustered together as in the WAIS these two tasks are part one subtest under the working memory index. The finding

that the IVR Digit Span tasks loaded on the same factor as the WAIS-IV Digit Span tasks indicates that the tasks are tapping into the same core latent abilities.

Discussion

The majority of participants who were administered the IVR tasks were able to complete all three tasks and only a small number of them (1%) had difficulties that caused them to complete only one or two of the tasks. Thus, it appears that the IVR system was easy to use and instructions were presented in a way that was understood by participants. Because there were no exclusion criteria except age, language proficiency, and self-exclusion from the study, the sample included people with lower IQ and a number of people with significant memory problems that would fit the description of mild cognitive impairment. Although we made no effort to identify people with mild dementia, anecdotal evidence indicated that a few of our participants had early dementia symptoms but were able to complete all three tests.

The correlation between the IVR scoring of the verbal fluency task and the clinician scoring was high ($r = .89$), and the verbal recognition error of the system was low (8% recognition error). However, perfect agreement between the automated scoring and the clinician scoring was found in only 41.1% of cases. In 53.2% of cases, the system failed to recognize responses that were part of the initial list and assigned lower scores to participants. Most of these voice recognition issues could be addressed by improvements in the speech recognition engines and programming the IVR system to detect behaviours such as low voice volume and speaking in sentences.

The mean number of fruits named by participants in this study tended to be lower ($M = 10.68$) compared to other studies using verbal fluency tasks (animals, fruits, items at the

supermarket) with similar samples; in these studies, the range of items produced ranged from 9 to 23 (Acevedo et al., 2000; Monsch et al., 1992; Tombaugh, Kozak, & Rees, 1999). This discrepancy could be due to the fact that previous studies had more stringent exclusion criteria and/or different recruitment strategies, which may have led to healthier and/or more educated participants.

One important limitation of our study was the lack of an in-person administration of an alternate version of the verbal fluency task. Thus, we compared participants' performance on the IVR verbal fluency task to their performance on the Similarities subtest of the WAIS-IV, which was the closest test in terms of cognitive abilities required (executive function and verbal abilities). There was only a modest relationship between the two variables ($r = .47$). However, the verbal fluency task correlated significantly with all verbal tests of the WAIS-IV providing support for the strong verbal component of the task.

The Digit Span forward and Digit Span backward administered by the system had very high correlations between the system and the clinician scoring ($r = .89$ and $r = .95$, respectively) and low verbal recognition error rate (5% and 7%, respectively); however, the scoring between the clinician and the system matched in only 63.8% of cases for the Digit Span forward task and in only 68% of cases for the Digit Span backward task. This was due to the same problems noted above for the verbal fluency task.

The correlations between the in-person administration of the Digit Span forward and backward and the IVR digit span tasks for the participants for which the system and the clinician scoring was identical were also modest ($r = .41$ for the Digit Span forward, and $r = .52$ for the Digit Span backward). The modest correlations indicate that a number of participants who obtained high scores on the IVR tasks received a lower score on the in-

person administration and vice versa (the scatterplot was oval.) We also noted that the mean performance of participants as a group was higher for the in-person administration compared to the IVR tasks. These results are consistent with another investigation, which reported higher mean scores for the in-person administration of tests compared to telephone administration (even when the administration over the phone is accomplished by a person) (Thompson, Prince, Macdonald, & Sham, 2001). In addition, Thompson et al., 2001 reported a number of low and moderate correlations between some of the tests for the in-person and telephone administration. Similarly, another study reported moderate correlation between the telephone and in-person administration of verbal learning tasks during a long delay condition (Unverzagt et al., 2007). However, in Unverzagt's study, the correlation between the Digit Span total score for telephone and in-person administration of the test was high ($r = .82$). The moderate correlation of the Digit Span scores in our study was somewhat surprising given that previous studies have reported comparable scores and considering that although the IVR Digit Span tasks were adapted from the WAIS-III and the in-person Digit Span tasks was from the WAIS-IV, they are essentially the same tasks with very similar administration and scoring. It is possible that some practice effects were in play for the in-person administration of the task, since the IVR was administered before the WAIS-IV. Thus, participants had some familiarity with the Digit Span forward and backward during the in-person administration. It is also possible that the IVR tasks are more demanding and require higher concentration due to the lack of non-verbal cues to mark the end of the digits presentation. For this reason, we introduced a beep at the end of each string for the IVR Digit Span tasks. However, the beep may have also introduced a distraction and may have interfered with participants' attention.

Lastly, our findings regarding the differences in correlations between the in-person administered Wechsler subtests and the IVR administered Digit Span tasks suggest that the different mode of administration of the two tasks may change what the tests measures. For example, the IVR tasks may be tapping on more than just the ability to hold numbers in mind and repeat them or hold information in mind manipulate it and generate a response.

In summary, the clinician and the IVR system scoring were highly correlated but we uncovered a number of issues related to both the administration and the scoring of the tests. Test results of the IVR administration of the tasks were not entirely comparable to the in-person administration. Some problems were obvious but there are some indications that IVR administration introduced new variables in known tests. Voice recognition software has significantly improved since our study was performed and the IVR software could take advantage of these improvements to control the administration and scoring issues we reported. However, we agree with the position expressed by Bauer (Bauer et al.2012), who cautioned that tests adapted for computer administration need to be validated and normed separately because the adaptation of the tests to a telephone interview may change them significantly.

References

- Acevedo, A., Loewenstein, D. A., Barker, W. W., Harwood, D. G., Luis, C., Bravo, M., Duara, R. (2000). Category fluency test: Normative data for english- and spanish-speaking elderly. *Journal of the International Neuropsychological Society*, 6(7), 760-769.
- Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M. , & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the american academy of clinical neuropsychology and the national academy of neuropsychology. *Archives of Clinical Neuropsychology*, 27, 362-373.
- Brandt, J, Spencer, M, & Folstein, M. (1988). The telephone interview for cognitive status. *Neuropsychiatry Neuropsychology Behavioral Neurology*, 1, 111-117.
- Brucki, S. M., & Rocha, M. S. (2004). Category fluency test: Effects of age, gender and education on total scores, clustering and switching in brazilian portuguese-speaking subjects. *Brazilian Journal of Medical and Biological Research*, 37(12), 1771-1777.
- Buckwalter, J. G., Crooks, V. C., & Petitti, D. B. (2002). A preliminary psychometric analysis of a computer-assisted administration of the telephone interview of cognitive status-modified. *Journal of Clinical and Experimental Neuropsychology*, 24(2), 168-175.
- Crooks, V. C., Petitti, D. B., Robins, S. B., & Buckwalter, J. G. (2006). Cognitive domains associated with performance on the telephone interview for cognitive status-modified. *American Journal of Alzheimer's Disease and Other Dementias*, 21(1), 45-53.
- de Jager, C. A., Budge, M. M., & Clarke, R. (2003). Utility of tics-m for the assessment of cognitive function in older adults. *International Journal of Geriatric Psychiatry*, 18(4), 318-324.
- Knopman, D. S., Knudson, D., Yoes, M. E., & Weiss, D. J. (2000). Development and standardization of a new telephonic cognitive screening test: The minnesota cognitive acuity screen (mcas). *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 13(4), 286-296.
- Knopman, D. S., Roberts, R. O., Geda, Y. E., Pankratz, V. S., Christianson, T. J., Petersen, R. C., & Rocca, W. A. (2010). Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. *Neuroepidemiology*, 34(1), 34-42.
- Lezak, M. D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological Assessment, 4th edition*. New York: Oxford University Press.

- Lucas, J. A., Ivnik, R. J., Smith, G. E., Bohac, D. L., Tangalos, E. G., Graff-Radford, N. R., & Petersen, R. C. (1998). Mayo's older americans normative studies: Category fluency norms. *Journal of Clinical and Experimental Neuropsychology*, *20*(2), 194-200.
- Miller, D.I., Bruce, H., Gagnon, M., Talbot, V., & Messier, C. (2011). Improving older adults' experience with interactive voice response systems. *Telemedicine Journal and E-Health*, *17*(6), 452-455. doi: 10.1089/tmj.2010.0204
- Monsch, A. U., Bondi, M. W., Butters, N., Salmon, D. P., Katzman, R., & Thal, L. J. (1992). Comparisons of verbal fluency tasks in the detection of dementia of the alzheimer type. *Archives of Neurology*, *49*(12), 1253-1258.
- Mundt, J. C., Ferber, K. L., Rizzo, M., & Greist, J. H. (2001). Computer-automated dementia screening using a touch-tone telephone. *Archives of Internal Medicine*, *161*(20), 2481-2487.
- Norton, M. C., Tschanz, J. A., Fan, X., Plassman, B. L., Welsh-Bohmer, K. A., West, N., Breitner, J. C. (1999). Telephone adaptation of the modified mini-mental state exam (3ms). The cache county study. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *12*(4), 270-276.
- Plassman, B L, Newman, T. T., Welsh, K. A., Helms, M., & Breitner, J. C. S. (1994). Properties of the telephone interview for cognitive status: Application in epidemiological and longitudinal studies. *Neuropsychiatry Neuropsychol Behavioral Neurology*, *7*, 235-241.
- Prince, M.J., MacDonald, A. M., Sham, P.C., Richards, M., Quraishi, S., & Horn, I. (1999). The development and initial validation of a telephone-administered cognitive test battery (tact). *International Journal of Methods in Psychiatric Research*, *8*(1), 49-57.
- Spreen, O., & Strauss, E. (1998). *A Compendium of Neuropsychological Tests*. New York: Oxford University Press.
- Thompson, Neil R., Prince, Martin J., Macdonald, Alison, & Sham, Pak C. (2001). Reliability of a telephone-administered cognitive test battery (tact) between telephone and face-to-face administration. *International Journal of Methods in Psychiatric Research*, *10*(1), 22-28. doi: 10.1002/mpr.97
- Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: Fas and animal naming. *Archives of Clinical Neuropsychology*, *14*(2), 167-177.
- Unverzagt, F. W., Monahan, P. O., Moser, L. R., Zhao, Q., Carpenter, J. S., Sledge, G. W., & Champion, V. L. (2007). The indiana university telephone-based assessment of

neuropsychological status: A new method for large scale neuropsychological assessment. *Journal of the International Neuropsychological Society*, 13(5), 799-806.

Wechsler, D. (1997). *Wechsler adult intelligence scale-3rd edition*. New York: Psychological Corporation.

Wechsler, D. (2008). *Wechsler adult intelligence scale, 4th edition*. San Antonio, TX: Pearson.

Welsh, K. A., Breitner, J. C. S., & Magruder-Habib, K. M. (1993). Detection of dementia in the elderly using telephone screening of cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioural Neurology* 6(2), 103-110.

Table 1.

Number of People With a Discrepancy Between Clinician Scoring and IVR Scoring

Discrepancy*	Frequency	Percent
-1.00	9	5.7
.00	65	41.1
1.00	32	20.3
2.00	27	17.1
3.00	13	8.2
4.00	5	3.2
5.00	4	2.5
7.00	2	1.3
8.00	1	.6
Total	158	100.0

* Positive numbers represent higher scores given by the clinician

Table 2.

Number of People With a Discrepancy Between IVR Scoring and Clinician Scoring for the Digit Span Forward Task

Discrepancy*	Frequency	Percent
-4.00	1	.7
-1.00	4	2.6
.00	97	63.8
1.00	30	19.7
2.00	14	9.2
3.00	5	3.3
4.00	1	.7
Total	152	100.0

* Positive numbers represent higher scores given by the clinician

Table 3.

Problems Causing Verbal Recognition Issues by the IVR (Digit Span Forward Task)

Problem	N	Percentage
<i>Overall discrepancy between scoring</i>	55	36.2
Participants correcting their responses	19	12.5
System did not allow for enough time to respond	6	3.9
System did not “hear” responses	14	9.2
Participants speaking in low voice	9	5.9
Foreign accent or bad pronunciation	21	13.8
Participant did not wait for the beep	20	13.2
Other verbal problems (e.g. coughing & “Ummm”)	15	9.9
Speaking in sentences	13	8.6

Table 4.

Number of Problems Encountered During the IVR Digit Span Forward Task

Number of Problems	Frequency	Percent
.00	98	64.5
1.00	17	11.2
2.00	15	9.9
3.00	19	12.5
4.00	2	1.3
5.00	1	.7
Total	152	100.0

Table 5.

Number of People With Discrepancy Between IVR Scoring and Clinician Scoring for the IVR Digit Span Backward Task

Discrepancy	Frequency	Percent
-2.00	1	.7
.00	104	68.0
1.00	31	20.3
2.00	11	7.2
3.00	6	3.9
Total	153	100.0

* Positive numbers represent higher scores given by the clinician

Table 6.

Problems Causing Verbal Recognition Issues by the IVR (Digit Span Backward Task)

Problem	N	Percentage
<i>Overall discrepancy between scoring</i>	49	32
Participants correcting their responses	15	9.8
System did not allow for enough time to respond	26	17
System did not “hear” responses	3	2
Participants speaking in low voice	6	3.9
Foreign accent or bad pronunciation	12	7.8
Participant did not wait for the beep	14	9.2
Other verbal problems (e.g. coughing & Ummm)	10	6.5
Speaking in sentences	10	6.5

Table 7.

Number of Problems Encountered During the IVR Digit Span Backward Task

Number of Issues	Frequency	Percent
.00	105	68.6
1.00	11	7.2
2.00	19	12.4
3.00	10	6.5
4.00	6	3.9
5.00	2	1.3
Total	153	100.0

Table 8.

Correlations Between the IVR Tests (for the People with Identical Clinician and IVR Scores) and the In-Person Administered WAIS-IV Battery.

	IVR Fluency	Similar.	IVR DSF	DSF	IVR DSB	DSB
Block D.	.096	.373**	-.188	.104	.196*	.165*
Similarities	.470**		.118	.160*	.206*	.160*
Digit Forward	-.121	.160*	.407**		.322**	.424**
Digit Backward	.015	.160*	.202*	.424**	.524**	
Matrix	.232	.438**	-.096	.105	.074	.228**
Vocabulary	.532**	.631**	.183	.261**	.232*	.253**
Arithmetic	-.071	.348**	.028	.287**	.452**	.333**
Symbol S.	-.019	.354**	.056	.176*	.315**	.216**
Visual Puz.	.134	.325**	.009	.133	.118	.241**
Information	.422**	.473**	.207*	.068	.046	.104
Coding	.014	.368**	.081	.163	.311**	.161

* Significant at the .05 level

** Significant at the .01 level

Table 9.

Principal Component Analyses With a Varimax Rotation for IVR Tasks and the Raw Scores of the 10 Core Subtests of the WAIS-IV.

Varimax rotated matrix			
	Components		
	1	2	3
Number of words without repetitions and errors	.036	.573	.186
IVR-COG digit Span Forward	-.185	.268	.697
IVR-COG digit Span Backward	.280	-.036	.684
Raw Score for Block Design	.715	.248	-.048
Raw Score for Similarities	.334	.718	.096
Raw Score for Digit Span Forward	.040	.133	.739
Raw Score for Digit Span Backward	.214	.009	.748
Raw Score for Matrix Reasoning	.627	.405	-.032
Raw Score for Vocabulary	.277	.771	.203
Raw Score for Arithmetic	.471	.220	.371
Raw Score for Symbol Search	.718	.062	.160
Raw Score for Visual Puzzles	.705	.141	.127
Raw Score for Information	.210	.778	-.075
Raw Score for Coding	.746	.150	.122

General Discussion

The popularity of IVR systems is continuing to grow. Today, it is virtually impossible to call an institution or a company and be greeted by a person. Instead, we hear an automated voice providing us with a list of choices from which we select options. Businesses were the first to adopt IVR technology after the cost savings related to the use of the technology was realized. However, other sectors such as healthcare have also used IVR. Many of the health-related applications are not only cost effective but also provide services that otherwise would not be provided due to lack of resources—reminders for medication refill, promotion of preventive care (i.e., screening and vaccination), and follow-up of patients after hospital discharge are examples of such services. Despite the growing popularity of IVR, research regarding the acceptability and utility of the technology, is still limited. For example very little is known about how older adults fare with these systems and what factors predict their attitudes, adoption and successful use of the technology. We attempted to fill this gap by examining the cognitive factors related to successful use of IVR in older people and evaluated their attitudes regarding IVR in general and as it relates to their success with four commercially available IVR applications. We also provided a proof-of-principle of using IVR for the purpose of administration and scoring of neuropsychological tests and discussed the benefits and limitations of the technology in that context.

In line with previous research (Katz et al., 1997; Settle et al., 1999), the results of our focus groups revealed negative attitudes towards IVR in older adults. The complaints that participants voiced were similar to the ones found in previous research. However, we also uncovered additional problems such as not being able to recover from mistakes, asking to wait too long on hold to speak to a person, being unable to reach an operator, and absence of

shortcuts in the system, which makes the interaction tedious. In our sample, a characteristics of a good IVR system included friendly, polite, with short, clear and unambiguous instructions. Participants also expressed a strong preference for being informed that they are speaking to a “computer” as many of them reported embarrassing experiences related to “forgetting” that they were interacting with an automated system and reverting to natural language. We uncovered the same problem in our subsequent study in which participants drifted from the short statements required by the IVR to full sentences when communicating with the voice-activated system. This is an interesting new finding, which may be related to familiarity with technology in general, working memory problems, or both.

In terms of the cognitive abilities required for successful IVR interaction, we found that working memory and auditory memory were strong predictors of older people’s ability to successfully complete IVR tasks. Age accounted for only 5% of the variance in older adults’ ability to cope with IVR. However, the oldest individuals in our sample were more likely to fail on all four tasks and committed more errors during their interaction. The inability to recover from errors may in fact be the reason why the oldest people in our sample were unable to complete any of the tasks. These findings are also in line with previous research (Dulude, 2002), which found that older adults were more likely to make a wrong selection and had more difficulties recovering from errors which frequently led to the discontinuation of the task. This is likely due to interplay between older adults cognitive difficulties and shortcoming in IVR design, which often does not provide recovery from error options (i.e., repeat and back options) at each level of the interaction. This was further supported by our findings that a large number of our sample did not use the repeat and back options.

By and large, our participants had significant difficulties interacting with the four IVR systems. In fact, only 3.2% of them were able to complete all four tasks, and 20.5% of them were unable to cope with any of the systems. This is an alarming finding, considering that the applications used in our study were commercially available IVR, with assumed universal design. The one system, Service Canada, which provides pension information, and is therefore geared towards older customers, had a success rate of only 50.2% in our sample. This indicated that only half of the people for whom the system was intended were able to benefit from its services. Statistics Canada IVR had a similar rate of success (57.8%), and the two airline systems were even more difficult for participants with 21.6% of success rate for United Airlines IVR and 24.9% for Air Canada IVR.

Not surprisingly, we found that participants' overall ratings of the systems in terms of their usability were related to their success with the IVR systems. The two systems with higher success rate (Service Canada and Statistics Canada) were rated more favourably compared to the United Airlines and Air Canada IVR. Nevertheless, a large number of our sample (50.8%), who were able to complete only one or two tasks, gave high overall ratings for the IVR regardless of their difficulties in interacting with the systems. These participants were not different from the rest of our sample in terms of age, education and self-reported memory status. The only demographic variable, which accounted for the differences between the two groups, was full-scale IQ: participants who gave higher ratings despite their relative lack of success had lower IQ scores. This finding has important implications for IVR design and for beta testing of IVR applications.

Lastly, we reported on a proof-of-principle of using voice-activated IVR for the purpose of administration and scoring of neuropsychological tests in a sample of participants

that was representative of the population of older adults aged 65 and older in Canada in terms of their demographics and full-scale IQ (FSIQ ranged from 68 to 134). We found that the majority of participants (99.96%) were able to complete all three IVR administered tests. This shows that it is possible to provide instructions for neuropsychological tests and administer these tests successfully using IVR. Small modifications to standard instructions were made (including a beep at the end of each digit string to indicate the end of the presentation) due to lack of visual cues when using the telephone as a medium of communication. We found a high correlation ($< .80$) between the IVR scoring of the tests and the scoring completed by a clinician and the verbal recognition error of the system was very low. Nevertheless, the IVR's failure to recognize some responses frequently led to an erroneous early discontinuation of two of the tasks and this was reflected by the generally lower scores obtained on the IVR administration of the tests compared to the in-person administration. There was also an indication that the different mode of administration (IVR vs. in-person) may have led to differences in the constructs that the tests measured. This was possibly due to higher auditory and working memory demand when using IVR as a medium.

Limitations

One important limitation of our research was that we did not measure hearing loss in our sample. Although, participants were allowed to adjust the volume of the phone and all of them were able to follow and complete the study protocol without notable hearing problems, it is possible that subtle changes in hearing abilities may have directly or indirectly affected their performance by increasing the cognitive load required to complete the IVR tasks. Future studies should include hearing evaluation when examining older adults' ability to interact with the technology.

In the literature, anxiety toward computer use and familiarity with the technology have been found to impact people's ability to successfully perform computer and web-based tasks (Czaja, et al., 2006; Matanda, Jenvey, & Phillips, 2004). These were not evaluated in our study; thus, the importance of anxiety and previous experience with the technology for IVR use remain to be examined in future research.

Although, we included two batteries of cognitive tests measuring intellectual and memory abilities, other cognitive factors such as prospective memory and executive functions were not evaluated. Considering the wide variety of IVR applications and the relative lack of uniformity in designing IVR systems, it is conceivable that previous experience with other applications would not easily generalize to a new system. Thus, each system would present as a novel task and successful interaction may depend on executive functions such as problem solving, planning and ability to cope with novelty. In addition, prospective memory may be involved in remembering to remember the task's requirements in each subsequent step of the interaction. Future research is needed to evaluate the importance of executive functions and prospective memory for successful communication with IVR systems.

We made an effort to collect a sample that is representative of the population and included a number of participants with lower socio-economic status and lower level of education. Our only exclusion criteria was age younger than 65 and non proficiency in English. It is possible that our sample also included older adults with early cognitive decline, such as MCI. Comparison between people with MCI and healthy older adults in ability to successfully cope with IVR tasks was not the focus of our study; therefore, we did not screen participants for the presence of MCI. However, this may be an important venue for future

investigations, specifically in regards to health-related IVR that targets older adults with cognitive decline and dementia.

In addition, because our focus was on examining the cognitive factors predicting IVR performance in older adult population, we did not include a sample of younger adults in our studies. It would be important to know if the same cognitive abilities also predict successful communication with IVR technology in younger adults. Such finding may provide an argument in favour of universal design that takes into account the heavy demands on working memory and auditory memory when interacting with IVR in both younger and older adults. Future studies, which include a sample of younger and older adults, may be able to answer that question.

Conclusions

- 1) The generally negative attitudes of older people towards IVR systems are mainly related to the poor design of most application and their lack of adaptability to older users' limitations. However, older adults are willing to use the technology and appreciate the value of IVR applications in providing health-related services.
- 2) IVR design should focus on simpler menus with fewer levels and provide error recovery options at each level of the interaction to optimise success and reduce frustration.
- 3) We demonstrated that working memory and auditory memory are predictive of older adults' ability to successfully complete IVR tasks. Thus, applications geared towards older users should pay special attention to reducing working and auditory memory demands. Alternatively, IVRs could be designed in a way that allows for adaptation to users' abilities and automatic switching between low or high memory loads.

- 4) Administration and scoring of neuropsychological tests using IVR is a feasible alternative to an in-person assessment. The new generation voice-recognition engines would likely be able to improve the reliability of voice-activated IVR test administration and scoring.

References

- Anderson, C., Soderpalm Gordth, A. H. V., & Berglund, M. (2007). Use of real-time Interactive Voice Response in a study of stress and alcohol consumption. *Alcoholism: Clinical and Experimental Research, 31*(11), 1908-1912.
- Anstey, K. J., & Low, L. F. (2004). Normal cognitive changes in aging. *Aust Fam Physician, 33*(10), 783-787.
- Bardone, A. M., Krahn, D. D., Goodman, B. M., & Searles, J. S. (2000). Using interactive voice response technology and timeline follow-back methodology in studying binge eating and drinking behavior: different answers to different forms of the same question? *Addict Behav, 25*(1), 1-11.
- Birdi, K., & Zapf, D. (1997). Age differences in reactions to errors in computer-based work. *Behaviour & Information Technology, 16*, 309-319.
- Corkrey, R., Parkinson, L., & Bates, L. (2005). Pressing the key pad: trial of a novel approach to health promotion advice. *Prev Med, 41*(2), 657-666.
- Crawford, A. G., Sikirica, V., Goldfarb, N., Popiel, R. G., Patel, M., Wang, C., et al. (2005). Interactive voice response reminder effects on preventive service utilization. *Am J Med Qual, 20*(6), 329-336.
- Czaja, S.J., Charness N., Fisk, A.D., Hertzog, C., Nair, S.N., Rogers, W.A., & Sharit, J. (2006). Factors predicting the use of technology: Findings from the Centre for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and Aging, 21*, 333-352.
- DeFrank, J., Rimer, B., Gierisch, J., Bowling, J. M., Farrel, D., & Skinner, C. (2009). Impact of mailed and automated telephone reminders on receipt of repeat mammograms: A randomized control trial. *Journal of Preventive Medicine, 36*(6), 459-467.
- Dulude, L. (2000). Usability of interactive voice response systems in real-life tasks performed by old and young women. *Unpublished manuscript*.
- Dulude, L. (2002). Automated telephone answering systems and aging. *Behaviour & Information Technology, 21*(3), 171-184.
- Dyck, J. L., & Smither, J. A. (1999). Age differences in computer anxiety: the role of computer experience, gender and education. *Journal of Educational Computing Research, 10*, 239-247.

- Estabrooks, P. A., & Smith-Ray, R. L. (2008). Piloting a behavioral intervention delivered through interactive voice response telephone messages to promote weight loss in a pre-diabetic population. *Patient Educ Couns*, 72(1), 34-41.
- Forster, A. J., & van Walraven, C. (2007a). Using an interactive voice response system to improve patient safety following hospital discharge. *Journal of Evaluation in Clinical Practice*, 13, 346-351.
- Forster, A. J., & van Walraven, C. (2007b). Using an interactive voice response system to improve patient safety following hospital discharge. *J Eval Clin Pract*, 13(3), 346-351.
- Gazzaley, A., & D'Esposito, M. (2007). Top-down modulation and normal aging. *Ann N Y Acad Sci*, 1097, 67-83.
- Gazzaley, A., Sheridan, M. A., Cooney, J. W., & D'Esposito, M. (2007). Age-related deficits in component processes of working memory. *Neuropsychology*, 21(5), 532-539.
- Graham, J., Tomcavage, J., Salek, D., Sciandra, J., Davis, D. E., & Stewart, W. F. (2012). Postdischarge monitoring using interactive voice response system reduces 30-day readmission rates in a case-managed Medicare population. *Med Care*, 50(1), 50-57.
- Greist, J. H., Marks, I. M., Baer, L., Kobak, K. A., Wenzel, K. W., Hirsch, M. J., et al. (2002). Behavior therapy for obsessive-compulsive disorder guided by a computer or by a clinician compared with relaxation as a control. *J Clin Psychiatry*, 63(2), 138-145.
- Greist, J. H., Mundt, J. C., & Kobak, K. (2002). Factors contributing to failed trials of new agents: can technology prevent some problems? *J Clin Psychiatry*, 63 Suppl 2, 8-13.
- Heisler, M., Halasyamani, L., Resnicow, K., Neaton, M., Shanahan, J., Brown, S., et al. (2007). "I am not alone": the feasibility and acceptability of interactive voice response-facilitated telephone peer support among older adults with heart failure. *Congest Heart Fail*, 13(3), 149-157.
- Jay, G. M., & Willis, S. L. (1992). Influence of direct computer experience on older adults' attitudes toward computers. *J Gerontol*, 47(4), P250-257.
- Katz, J. (1994). Empirical and theoretical dimensions of obscene phone calls to women in the United States. *Human Communications Research*(December), 155-180.
- Katz, M. J., Aspden, P., & Reich, W. A. (1997). Public attitudes toward voice-based electronic messaging technologies in the United States: A national survey of opinions about voice response units and telephone answering machines. *Behaviour & Information Technology*, 16(3), 125-144.

- Kim, H., Bracha, Y., & Tipnis, A. (2007). Automated depression screening in disadvantaged pregnant women in an urban obstetric clinic. *Arch Womens Ment Health, 10*(4), 163-169.
- Kim, H. G., Geppert, J., Quan, T., Bracha, Y., Lupo, V., & Cutts, D. B. (2012). Screening for postpartum depression among low-income mothers using an interactive voice response system. *Matern Child Health J, 16*(4), 921-928.
- Kobak, K. A., Greist, J. H., Jefferson, J. W., & Katzelnick, D. J. (1996). Computer-administered clinical rating scales. A review. *Psychopharmacology (Berl), 127*(4), 291-301.
- Lee, H., Friedman, M. E., Cukor, P., Ahem, D. (2003). Interactive voice response system (IVR) in health care services. *Nurs Outlook, 51*(6), 277-283.
- Lennig, M., Bielby, G., & Massicotte, J. (1995). Directory assistance automation in Bell Canada: Trail results. *Speech Communication, 17*, 227-234.
- Matanda, M., Jenvey, V. B., & Phillips, J. G. (2004). Internet use in adulthood: Loneliness, computer anxiety and education. *Behaviour Change, 21*(2), 103-114.
- Moore, H. K., Mundt, J. C., Modell, J. G., Rodrigues, H. E., DeBrotta, D. J., Jefferson, J. J., et al. (2006). An examination of 26,168 Hamilton Depression Rating Scale scores administered via interactive voice response across 17 randomized clinical trials. *J Clin Psychopharmacol, 26*(3), 321-324.
- Morrow, D., Leirer, V. O., Carver, L. M., Tanke, E. D., & McNally, A. D. (1999). Repetition improves older and younger adult memory for automated appointment messages. *Hum Factors, 41*(2), 194-204.
- Mundt, J. C., Bohn, M. J., King, M., & Hartley, M. T. (2002). Automating standard alcohol use assessment instruments via interactive voice response technology. *Alcohol Clin Exp Res, 26*(2), 207-211.
- Mundt, J. C., Geralts, D. S., & Moore, H. K. (2006). Dial "T" for testing: Technological flexibility in neuropsychological assessment. *Telemed J E Health, 12*(3), 317-323.
- Mundt, J. C., Kinoshita, L. M., Hsu, S., Yesavage, J. A., & Greist, J. H. (2007). Telephonic Remote Evaluation of Neuropsychological Deficits (TREND): longitudinal monitoring of elderly community-dwelling volunteers using touch-tone telephones. *Alzheimer Dis Assoc Disord, 21*(3), 218-224.
- Nakagawa, A., Marks, I. M., P, J.M., Bachofen, M., Bear, L., Dottl, S.L., Greist, J.H. (2000). Treatment of obsessive-compulsive disorder guided by manual and computer-conducted telephone interview. *Journal of Telemedicine and Telecare, 6*, 22-26.

- Naylor, M. R., Keefe, F. J., Brigidi, B., Naud, S., & Helzer, J. E. (2008). Therapeutic Interactive Voice Response for chronic pain reduction and relapse prevention. *Pain, 134*(3), 335-345.
- Oake, N., van Walraven, C., Rodger, M. A., & Forster, A. J. (2009). Effect of an interactive voice response system on oral anticoagulant management. *Cmaj, 180*(9), 927-933.
- Osgood-Hynes, D. J., Greist, J. H., Marks, I. M., Baer, L., Heneman, S. W., Wenzel, K. W., et al. (1998). Self-administered psychotherapy for depression using a telephone-accessed computer system plus booklets: an open U. S.-U. K. study. *Journal of Clinical Psychiatry, 59*, 358-365.
- Park, H. L., O'Connell, J. E., & Thomson, R. G. (2003). A systematic review of cognitive decline in the general elderly population. *Int J Geriatr Psychiatry, 18*(12), 1121-1134.
- Possemato, K., Kaier, E., Wade, M., Lantinga, L. J., Maisto, S. A., & Ouimette, P. (2012). Assessing daily fluctuations in posttraumatic stress disorder symptoms and substance use with Interactive Voice Response technology: protocol compliance and reactions. *Psychol Serv, 9*(2), 185-196.
- Regan, S., Reyen, M., Lockhart, A. C., Richards, A. E., & Rigotti, N. A. (2011). An interactive voice response system to continue a hospital-based smoking cessation intervention after discharge. *Nicotine Tob Res, 13*(4), 255-260.
- Reidel, K., Tamblyn, R., Patel, V., & Huang, A. (2008). Pilot study of an interactive voice response system to improve medication refill compliance. *BMC Med Inform Decis Mak, 8*, 46.
- Rose, G. L., Skelly, J. M., Badger, G. J., Naylor, M. R., & Helzer, J. E. (2012). Interactive voice response for relapse prevention following cognitive-behavioral therapy for alcohol use disorders: a pilot study. *Psychol Serv, 9*(2), 174-184.
- Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S., Mundt, J. C., et al. (2006). An evaluation of the quick inventory of depressive symptomatology and the hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. *Biol Psychiatry, 59*(6), 493-501.
- Sano, M., Egelko, S., Ferris, S., Kaye, J., Hayes, T. L., Mundt, J. C., et al. (2010). Pilot study to show the feasibility of a multicenter trial of home-based assessment of people over 75 years old. *Alzheimer Dis Assoc Disord, 24*(3), 256-263.
- Schroder, K. E., Johnson, C. J., & Wiebe, J. S. (2007). Interactive Voice Response Technology applied to sexual behavior self-reports: a comparison of three methods. *AIDS Behav, 11*(2), 313-323.

- Settle, R. B., Dillon, T. W., & Alreck, P. L. (1999). Acceptance of phone-based interface for automated call direction. *Behaviour and Information Technology*, 18(2), 97-107.
- Sharit, J., Czaja, S. J., Nair, S., & Lee, C. C. (2003). Effects of age, speech rate, and environmental support in using telephone voice menu systems. *Hum Factors*, 45(2), 234-251.
- Sinadinovic, K., Wennberg, P., & Berman, A. H. (2011). Population screening of risky alcohol and drug use via Internet and Interactive Voice Response (IVR): a feasibility and psychometric study in a random sample. *Drug Alcohol Depend*, 114(1), 55-60.
- Smith, M. M., Tremont, G., & Ott, B. R. (2008). A Review of Telephone-Administered Screening Tests for Dementia Diagnosis. *Am J Alzheimers Dis Other Demen*.
- Stacy, J. N., Schwartz, S. M., Ershoff, D., & Shreve, M. S. (2009). Incorporating tailored interactive patient solutions using interactive voice response technology to improve statin adherence: results of a randomized clinical trial in a managed care setting. *Popul Health Manag*, 12(5), 241-254.
- Toll, B. A., Cooney, J. L., McKee, S. A., O'Malley, S. S., & Cooney, N. L. (2008). Correspondence of interactive voice response (IVR) reports of nicotine withdrawal, craving, and negative mood with questionnaire ratings. *Nicotine & Tobacco Research*, 10(6), 1057-1064.
- Treitz, F. H., Heyder, K., & Daum, I. (2007). Differential course of executive control changes during normal aging. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*, 14(4), 370-393.
- Tucker, J. A., Foushee, H. R., & Black, B. C. (2008). Behavioural economic analysis of natural resolution of drinking problems using IVR self-monitoring. *Experimental and Clinical Psychopharmacology*, 16(4), 332-340.
- Tucker, J. A., Roth, D. L., Huang, J., Crawford, M. S., & Simpson, C. A. (2012). Effects of interactive voice response self-monitoring on natural resolution of drinking problems: utilization and behavioral economic factors. *J Stud Alcohol Drugs*, 73(4), 686-698.
- Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: a review of meta-analyses. *Neurosci Biobehav Rev*, 26(7), 849-857.
- Westerman, S. J., Davies, D. R., Glendon, A. I., Stammers, R. B., & Matthews, G. (1995). Age and cognitive ability as predictors of computerized information retrieval. *Behaviour and Information Technology*, 14, 313-326

Appendices

Appendix A Advertisement for Focus Groups

Appendix B Advertisement for the Validation of the IVR-COG

Appendix C Description of the four IVR tasks

Appendix D Usability Questionnaire

Appendix E The Health Questionnaire

Appendix F Screening Form for Participant Selection for the Focus Groups and the Validation of the IVR-COG.

Appendix G Focus Groups Consent Form

Appendix H Results Disclosure Form

Appendix I Validation of the IVR-COG Consent Form

Appendix J Instructions to participants for the four IVR tasks

Appendix A

PEOPLE 65 YEARS AND OLDER NEEDED FOR A STUDY ON MEMORY

Dr. Claude Messier and his collaborators at the University of Ottawa and the Institute of Mental Health Research (Royal Ottawa Hospital) are looking for people aged 65 years and older to help develop new ways to test mental abilities using computerized interactive voice systems.

You are asked to participate in three focus group meetings where you will give us your opinion and ideas about different elements of a computerized voice response system that we want to develop. We want to adapt automatic phone services to seniors use to make them easier to use.

Volunteers for this important study should be:

- ◆ available for three 2 ½ hour visits
- ◆ able to come to the University of Ottawa for three visits

We are looking for 36 volunteers (18 men and 18 women) Of those we would like to have:

- * 12 people with no high school diploma
- * 12 people with a high school diploma
- * 12 people with a college or university diploma
- * at least 3 people aged 75 years and over

At the end of the study, you will receive \$100 for your participation.

Parking expenses, bus or taxi (if necessary) will be reimbursed
Refreshments will be served.

For more information, please contact:
Delyana Miller, Research Coordinator

Appendix B

PEOPLE 65 YEARS AND OLDER NEEDED FOR A STUDY ON MEMORY

Dr. Claude Messier and his collaborators at the University of Ottawa and the Institute of Mental Health Research (Royal Ottawa Hospital) are looking for people aged 65 years and older to help develop new ways to test mental abilities using computerized interactive voice systems.

You are asked to participate in one meeting where you will use a computerized voice response system that will provide a short test of your mental abilities. We will also test those same abilities using regular pen and pencil tests to verify the accuracy of the computerized test. Our goal is to adapt automatic phone services to seniors use to make them easier to use.

Volunteers for this important study should be:

- ◆ Available for two visits at the University of Ottawa (The first visit lasts 2 ½ hours, and the second visit- 30 min.)

We are looking for 210 volunteers (105 men and 105 women) Of those we would like to have:

- * 70 people with no high school diploma
- * 70 people with a high school diploma
- * 70 people with a college or university diploma
- * at least 52 people aged 75 years and over

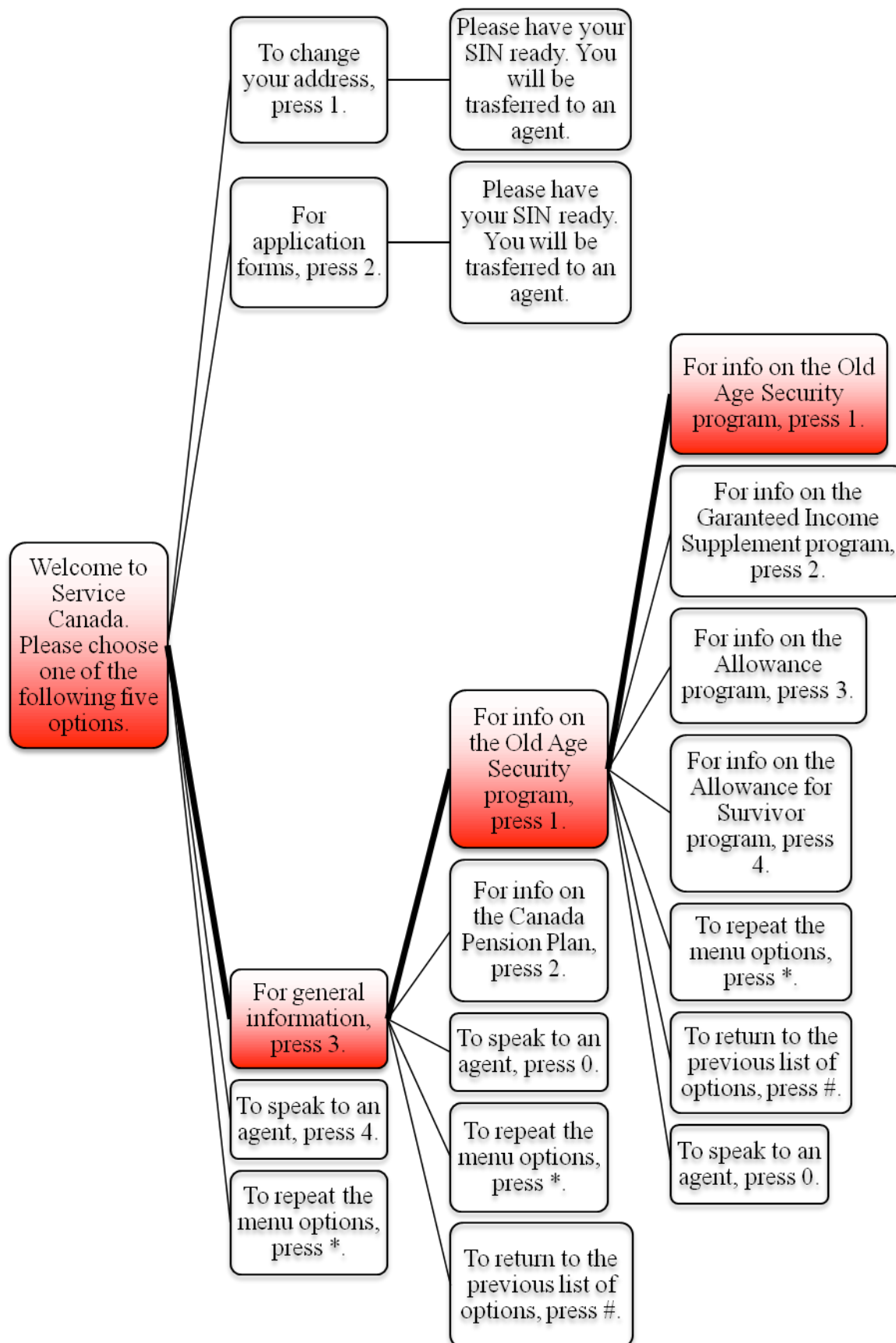
At the end of the study, you will receive \$100 for your participation.

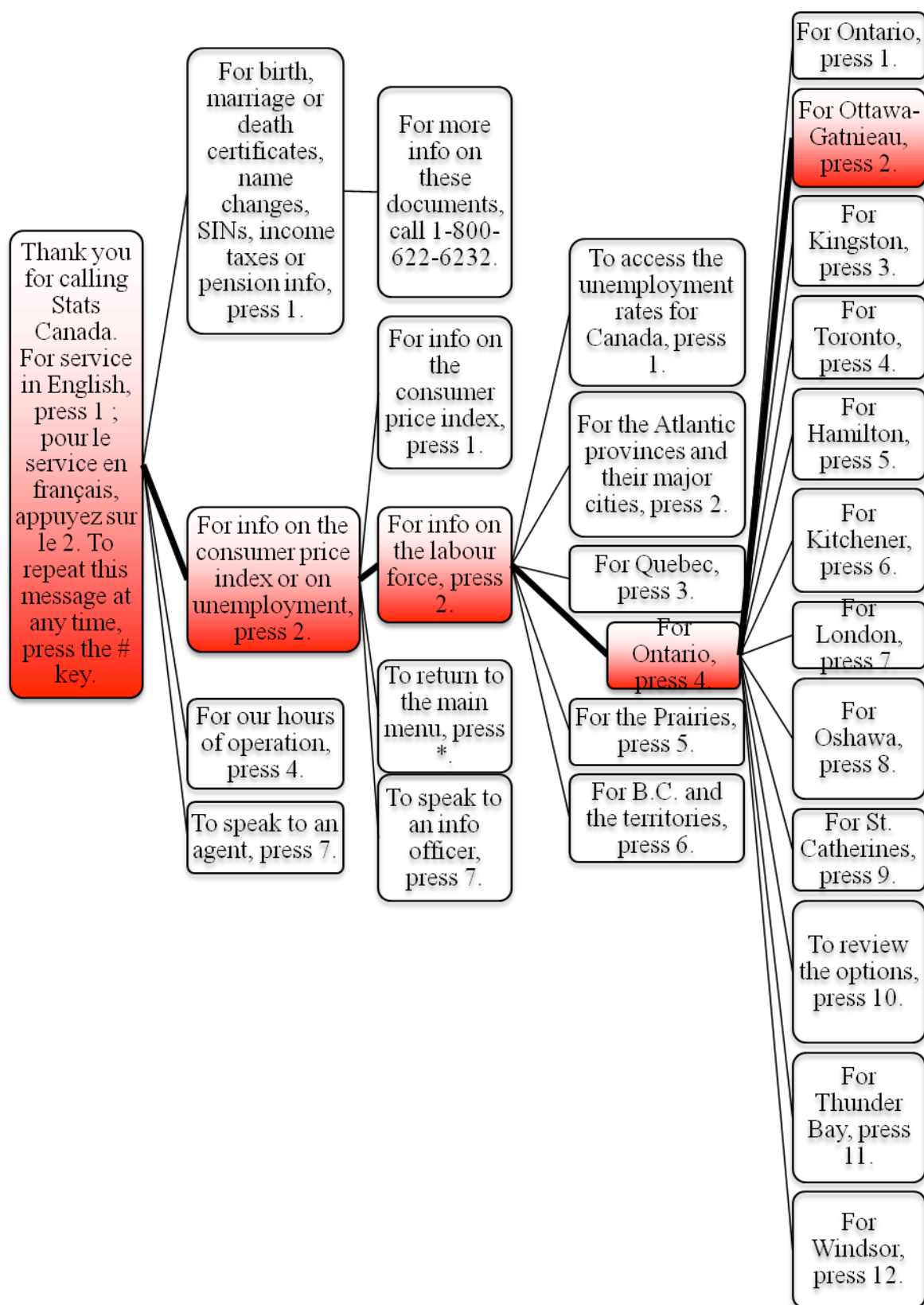
Parking expenses, bus or taxi (if necessary) will be reimbursed
Refreshments will be served.

For more information, please contact:
Delyana Miller, Research Coordinator

Appendix C

Description of the four IVR systems





Thank you for calling United Airlines



Please select one of the following four options:



For departure and arrival information – Reservations – Mileage Plus – More options



Enter or say the flight number, or say “**I don’t know it**”.



Do you want information on arrival or **departure**?



You can interrupt me or say “start over” at any time.



What is the departure city? **Toronto**



What is the arrival city? **New York**



United Airlines service has more than one airport in New York. They are: New York, New

Jersey, **J.F.K., New York**, and La Guardia, New York.



The scheduled departure time? **6 p.m.**



Let’s make sure I got that right. (Provided information is repeated; however, the date is today’s

date). Is that right? Yes or **No**



Which one do you want to change? Departure city, arrival city, **date**, or time?



What date would you like? **Tomorrow**



I found two that you might want. Let’s find the right one.



Select flight number and departure time

Thank you for calling Air Canada



. For service in English **press 1**, pour continuer en français, faites le 2.



Please choose one of the following six options:

For flight arrival and departure information, press 1

To search for lowest fares, press 2 – For frequently asked questions, press 3

For reservations, press 4 – For assistance with Air Canada.com, press 5

For assistance with flight passes, press 6

To repeat this message, press 9.



If you know the flight number, press 1 – **If you don't know the flight number, press 2**



If you would like to hear the instructions on how to use the flight information system, press



Otherwise, if you have a touch tone phone, press 2.

If at any time you need assistance or would like to start over, press *.



To hear arrival and departure information for today, press 1 – **For tomorrow, press 2**

For any other day, press 3.



Please enter the first 4 letters of the departure city or enter the 3 letter airport code followed by

the pound key. For Q or Z, press the 0 key: **TORO#** or **YYZ#**



For Toronto Island, press 1 – **For Toronto Pearson, press 2**



Please enter the first 4 letters of the arrival city or the 3 letter airport code followed by the pound

key. For Q or Z, press 0: **VANC#** or **YVR#**



Vancouver



To hear arrival times, press 1. – **To hear departure times, press 2.**



Enter the departure time followed by the pound key: **1200#**

↓
For A.M., press 1 – For PM, press 2

↓
Write down your flight information and departure time.

Appendix D

USABILITY QUESTIONNAIRE

(adapted form Dulude, 2000)

On the scale from one to five shown bellow, indicate if you agree or disagree with the ten statements on this page.

1	2	3	4	5
strongly agree	disagree	neutral	agree	strongly agree

Rating 1-5

11. I found the system service easy to use....._____
12. I felt very frustrated with this service....._____
13. I felt in control while using this service....._____
14. I thought some of the choices are confusing....._____
15. The operator's voice was very clear....._____
16. I am left with a good impression of that organization....._____
17. I had to concentrate really hard in doing this task....._____
18. I sometimes felt lost while using this service....._____
19. I enjoyed using this service....._____
20. I thought the operator spoke too quickly....._____

Complementary questions:

Had you ever used this system or a similar one before? How many times?

What was going on through your head when you were doing the task?

Do you have any other comments you would like to add about this task?

Appendix E

Health Questionnaire

What is your age ? Age has to be over 65 Age = _____ /_/_/1

How many years of schooling have you completed? _____ /_/_/2

Ethnicity _____

Yes = 1 No = 0

Medical history

Are you diabetic? _____ /_/_3

Do you suffer from important memory problems? _____ /_/_4

Did you previously have a hemorrhage (intracranial)? _____ /_/_5

Did you suffer from a brain tumor or lesion? _____ /_/_6

Did you ever had any brain disease? _____ /_/_7

Do you have chronic hepatitis? _____ /_/_8

Do you suffer from high blood pressure requiring medication? _____ /_/_9

Did you previously suffer from a loss of consciousness lasting more than 1 hour? _____ /_/_10

Are you seeing a psychiatrist presently? _____ /_/_11

Are you currently treated for depression ? _____ /_/_12

- Now I would like to ask you a few questions on your memory . **Yes=1 No=0**

1) Do you have memory problems? _____ /_/_13

2) Do you have difficulty remembering the names of persons that have been
introduced to you? _____ /_/_14

3) Do you misplace objects _____ /_/_15

4) Do you have difficulty remembering a list of things to buy or a list of things to do? _____ /_/_16

5) Do you have problems remembering telephone numbers or postal codes? _____ /_/_17

6) Does it often take you a long time to remember information _____ /_/_18

7) Does distraction (e.g. noise, another person talking) make it harder for you
to remember information? _____ /_/_19

8) If you are experiencing some memory problems, would you say,
they came suddenly or their appearance was over a long period?
(**gradual=1 sudden=0**) _____ /_/_20

Appendix F

Screening participants for focus groups

Name:

Female: Male:

Age:

English proficiency:

Level of Education:

Additional Courses (if applicable):

Previous Occupation:

Present Occupation (if applicable):

How did you hear about the study?

What time of the day will be most convenient to you to come to University?

Type of transportation:

Appendix G

Consent form**Validation of an interactive voice response (IVR) protocol to evaluate cognitive status in an older population**

Name of Researcher: Delyana Miller, School of Psychology, University of Ottawa.

Supervisors: Dr. Claude Messier, School of Psychology, University of Ottawa, Dr Elizabeth Kristjansson, School of Psychology, University of Ottawa, and Dr. Michèle Gagnon, Institute of Mental Health Research, Royal Ottawa Hospital.

Invitation to Participate: I am invited to participate in the abovementioned research study conducted by Delyana Miller under the supervision of Dr. Claude Messier, Dr. Elizabeth Kristjansson and Dr. Michèle Gagnon.

Purpose of the Study: I understand that the purpose of the study is to develop a short test to measure mental abilities to be administered over the telephone.

Participation: My participation will consist essentially of coming for 3 sessions at the University of Ottawa lasting approximately two hours and a half during which I will be asked to give my opinion on a sample of questions that we want to use in our test of cognitive abilities. I will also be asked to comment on the voices used to record the questions and the general way in which the questions will be asked

The sessions have been scheduled for (date) and will take place at the University of Ottawa, Vanier Building.

1) _____

2) _____

Risks: There no special risks associated with your participation to this project.

Benefits: My participation in this study will not benefit me directly in any way. My participation to this experiment will help develop tests of mental abilities that can be administered over the telephone.

Confidentiality: I have received assurance from the researcher that the information I will share will remain strictly confidential. I understand that the contents will be used only for research purposes and that my confidentiality will be protected by rendering all the information anonymous after I complete the experiment. I also agree not tell other people what the other participants said during the focus group so that, what they say remains confidential.

Anonymity: Anonymity will be protected in the following manner in that my name will not be associated with the recordings made during the focus group, and my personal information and comments will not be communicated in any way except as part of group averages.

Appendix H

Results disclosure form

I, _____, am asking Dr Claude Messier to send my tests results obtained during my participation to his study entitled « Validation of an interactive voice response (IVR) protocol to evaluate cognitive status in an older population». I understand that only a summary of the results will be sent by mail and that you can only send these results to a registered medical doctor or a registered psychologist.

Participant's signature: _____ Date:
(Signature) (Date)

Researcher's signature: _____ Date:
(Signature) (Date)

Appendix I

Consent Form Validation of the IVR-COG**Validation of an interactive voice response (IVR) protocol to evaluate cognitive status in an older population**

Name of Researcher: Delyana Miller, School of Psychology, University of Ottawa.

Supervisors: Dr. Claude Messier, School of Psychology, University of Ottawa, Dr Elizabeth Kristjansson, School of Psychology, University of Ottawa, and Dr. Michèle Gagnon, Institute of Mental Health Research, Royal Ottawa Hospital.

Invitation to Participate: I am invited to participate in the abovementioned research study conducted by Delyana Miller under the supervision of Dr. Claude Messier, Dr. Elizabeth Kristjansson and Dr. Michèle Gagnon.

Purpose of the Study: I understand that the purpose of the study is to develop and validate a short test to measure mental abilities to be administered over the telephone.

Participation: My participation will consist essentially of coming for 2 visits at the University of Ottawa lasting approximately two hours and a half during which I will undergo testing of my cognitive abilities. One of the tests will be administered over the phone by a computerized system and the examiner will administer the rest of the tests. To participate in this study, I must be proficient in English.

The sessions have been scheduled for (date) and will take place at the University of Ottawa, Vanier Building.

- 1) _____
- 2) _____

Risks: There no special risks associated with your participation to this project.

Benefits: My participation in this study will not benefit me directly in any way. My participation to this experiment will help develop tests of mental abilities that can be administered over the telephone.

Confidentiality: I have received assurance from the researcher that the information I will share will remain strictly confidential. I understand that the contents will be used only for research purposes and that rendering all the information anonymous after I complete the experiment will protect my confidentiality.

Anonymity: Anonymity will be protected in the following manner in that my results will remain anonymous and my personal information and results will not be communicated in any way except as part of group averages.

Retention of Data: The data collected (results at mental activity tests, answers to questionnaire) will be kept in a secure manner in a locked room at the University of Ottawa. The numeral key identifying my results will be kept in a locked cabinet during my participation to the study and will be destroyed at the end of my participation. The cabinet will be accessible only to the researcher and the supervisors named above.

Compensation: I will receive \$100.00 for my participation in this study. In addition, my parking expenses and reasonable travel expenses (bus or taxi if necessary) will be reimbursed.

Voluntary Participation: I am under no obligation to participate and if I choose to participate, I may withdraw from the study at any time and/or refuse to answer any questions. If I choose to withdraw, all data gathered until the time of withdrawal will be made anonymous but may be used by the researcher if needed.

Acceptance: I, (_____), agree to participate in the above research study conducted by (Delyana Miller) of the School of Psychology, University of Ottawa) under the supervision of Dr. Claude Messier, Dr Elizabeth Kristjansson and Dr. Michèle Gagnon. I understand that by accepting to participate I am in no way waiving my right to withdraw from the study.

If I have any questions about the study, I may contact the researcher or supervisors at the numbers mentioned above.

If I have any ethical concerns regarding my participation in this study, I may contact the Protocol Officer for Ethics in Research, University of Ottawa, 550 Cumberland Street, Room 159, Ottawa, ON K1N 6N5 (613) 562-5841 or ethics@uottawa.ca.

There are two copies of the consent form, one of which is mine to keep.

Participant's signature: _____ Date: _____
(Signature) (Date)

Researcher's signature: _____ Date: _____
(Signature) (Date)

Appendix J

Instructions to participants for the four IVR tasks

Instructions before beginning the IVR tasks: Now I will ask you to call four automated systems and obtain specific information from them. I will give you a list with the instructions for each task prior to the call. Please avoid using the options that will connect you with a live operator and use only the automated choices. If you hang up, the test will discontinue. You will be allowed to dial the system's number again only if you get a busy signal or if the initial number is incorrect. If you make a mistake while interacting with any of the four systems, please try to recover from it within the same call. (The examiner will be allowed to give additional instructions and explanations to participants and the task will not begin until it is clear that participants understand the task.)

Instruction to participants for Task 1: Call Service Canada at 1 800 277 9914 and obtain information on what you need to qualify for an Old Age Security Pension.

Instructions to participants for Task 2: Call Statistics Canada at 1 800 263 1136 and obtain the latest information on the unemployment rate for the Ottawa-Gatineau region.

Instructions to participants for Task 3: A friend who lives in New York wants to see you right away and offers to pay your way there. You decide to leave by plane from Toronto **tomorrow**. Call United Airlines at 1 800 864 8331 or 1 800 241 6522 and find out if there is a flight from Toronto to New York (Kennedy airport) around 6:00pm tomorrow evening. Note down the flight number and departure time.

Instructions to participants for Task 4: A friend who lives in Vancouver had invited you to visit him as soon as possible. You decide to fly from Toronto **tomorrow** at noon. Call Air Canada at 1 888 247 2262 and find out if there are flights from Toronto to Vancouver around 12:00 pm tomorrow. Note down the flight number and departure time. You will be asked to provide the three-letter code of the departure and arrival cities. Here are the codes for the two airports: Toronto- YYZ and Vancouver- YVR.