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

Delyana Miller,¹ Michèle Gagnon,¹ Vincent Talbot,² and Claude Messier¹

¹School of Psychology, University of Ottawa, Ontario K1N 6N5, Canada.
²TelAsk Technologies, Ottawa, Ontario K2C 3W7, Canada.

Objectives. Interactive voice response (IVR) systems 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 toward these systems have consistently found that people in general and older people in particular strongly dislike these systems. 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 4 real-life IVR systems that included fact-finding at governmental agencies and plane ticket reservation.

Results. The results indicated that adults aged 65 and older experience significant difficulties in interacting with IVR systems. A significant number of people (20.5%) could not complete any of the tasks. Participants who could not complete any task were 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 memory 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.

Key Words: Automated telephone systems—Cognitive abilities—Interactive voice response—Memory—Older people.

Interactive voice response (IVR) technology is a telephony 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 internet at home (Statistics Canada, 2009). There has been a growing interest in adopting IVR technology in the health care 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 health care domain in which IVR systems have been shown to be effective in increasing health behaviors. Several studies have shown the utility and feasibility of IVR systems in increasing the frequency of preventive care behaviors 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, Geralts, & 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 health care 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 Tests. 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 concentration, 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 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 were reviewed and approved by the Research Ethics Board of the University of Ottawa. One hundred and eighty-five (120 female participants: 65%) community-dwelling people between 65 and 92 years of age (mean = 73.32 years, standard deviation [SD] = 6.44) were recruited from diverse socioeconomic 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 degree, 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 hemorrhagic stroke, 1.1% had been treated for a brain tumor, 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 with 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) comprised 10 core subtests and 5 optional subtests, 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 took 67 min to complete the 10 core subtests.
The 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 took 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, whereas 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 6:00 p.m. 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 p.m. 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 four levels in order to complete the task. Statistics Canada 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 four), 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.

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 with 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 with 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 ($β$), the semipartial correlations ($sr^2$), and $R^2$; 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 $R^2 = .23, F(3, 181) = 24.86, p < .001$. 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

<table>
<thead>
<tr>
<th>Number of tasks completed</th>
<th>%</th>
<th>Age</th>
<th>Education</th>
<th>Full-scale IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20.5</td>
<td>$M = 76.29, SD = 7.54$</td>
<td>$M = 13.26, SD = 3.17$</td>
<td>$M = 92.14, SD = 10.84$</td>
</tr>
<tr>
<td>1</td>
<td>26.5</td>
<td>$M = 73.29, SD = 6.25$</td>
<td>$M = 13.65, SD = 2.46$</td>
<td>$M = 102, SD = 12.86$</td>
</tr>
<tr>
<td>2</td>
<td>31.9</td>
<td>$M = 72.83, SD = 6.26$</td>
<td>$M = 14.03, SD = 2.97$</td>
<td>$M = 105.39, SD = 13.55$</td>
</tr>
<tr>
<td>3</td>
<td>17.8</td>
<td>$M = 71.06, SD = 5.13$</td>
<td>$M = 14.67, SD = 3.15$</td>
<td>$M = 107.12, SD = 10.86$</td>
</tr>
<tr>
<td>4</td>
<td>3.2</td>
<td>$M = 72.17, SD = 2.6$</td>
<td>$M = 14, SD = 1.9$</td>
<td>$M = 115.33, SD = 9.85$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Range: 68–75</td>
<td>Range: 12–16</td>
<td>Range: 106–131</td>
</tr>
</tbody>
</table>

*Note. SD = standard deviation.*
variable is weighted more heavily for the two airplane tasks compared with the two simpler information tasks. Separate analyses that used levels attained for tasks 1 and 2 and levels attained for tasks 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 tasks 3 and 4, compared with the two simpler information tasks. Separate analyses that used levels attained for 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 logistic 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 $R^2 = .07, F(2, 182) = 8.39, p < .001$. The addition of working memory to the equation with age resulted in a small but significant increment in the adjusted $R^2$. Step 2, which added working memory in the equation, led to an adjusted $R^2 = .05, F(1, 183) = 11.14, p = .001$. Step 3, which added auditory memory in the equation together

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

Because there were larger number of levels in tasks 3 and 4 and thus higher chance for committing errors during the interaction with these two systems compared with 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 logistic 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 $R^2 = .07, F(2, 182) = 8.39, p < .001$. The addition of working memory to the equation with age resulted in a small but significant increment in the adjusted $R^2$. Step 2, which added working memory in the equation, led to an adjusted $R^2 = .05, F(1, 183) = 11.14, p = .001$. Step 3, which added auditory memory in the equation together

### Table 2. Sequential Regression of Age (Reciprocal), Cognitive Function, and Memory on Number of IVR Tasks Completed

<table>
<thead>
<tr>
<th>Variables</th>
<th>No. of IVR tasks</th>
<th>Age (reciprocal)</th>
<th>WM</th>
<th>AM</th>
<th>PR</th>
<th>VC</th>
<th>PS</th>
<th>VM</th>
<th>$B$</th>
<th>$SEB$</th>
<th>$\beta$</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>$sr^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.23</td>
<td>.17</td>
<td>.31</td>
<td>.36</td>
<td>.40</td>
<td>.46</td>
<td>.56</td>
<td>.40</td>
<td>.46</td>
<td>-.00</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>WM</td>
<td>.43</td>
<td>.17</td>
<td>.31</td>
<td>.36</td>
<td>.40</td>
<td>.46</td>
<td>.56</td>
<td>.40</td>
<td>.46</td>
<td>-.00</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>AM</td>
<td>.36</td>
<td>.32</td>
<td>.40</td>
<td>.28</td>
<td>.46</td>
<td>.46</td>
<td>.50</td>
<td>.50</td>
<td>.46</td>
<td>-.00</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>PR</td>
<td>.36</td>
<td>.32</td>
<td>.40</td>
<td>.28</td>
<td>.46</td>
<td>.46</td>
<td>.50</td>
<td>.50</td>
<td>.46</td>
<td>-.00</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>VC</td>
<td>.33</td>
<td>.45</td>
<td>.37</td>
<td>.35</td>
<td>.56</td>
<td>.40</td>
<td>.00</td>
<td>.04</td>
<td>.04</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>PS</td>
<td>.31</td>
<td>.44</td>
<td>.31</td>
<td>.47</td>
<td>.50</td>
<td>.54</td>
<td>.34</td>
<td>.34</td>
<td>.46</td>
<td>-.00</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.85</td>
<td>-2.85</td>
<td>.46</td>
<td>.70</td>
<td>.55</td>
<td>.29</td>
<td>.23</td>
<td>.23</td>
<td>.23</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>SD</td>
<td>1.57</td>
<td>.01</td>
<td>40.56</td>
<td>81.28</td>
<td>55.24</td>
<td>77.14</td>
<td>78.14</td>
<td>44.43</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.31</td>
<td>.31</td>
<td>.40</td>
<td>.28</td>
<td>.46</td>
<td>.46</td>
<td>.50</td>
<td>.50</td>
<td>.46</td>
<td>-.00</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

**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. AM = auditory memory; B = unstandardized regression coefficient; $\beta$ = standardized regression coefficient; PR = perceptual reasoning; PS = processing speed; SD = standard deviation; SE B = intercept; $sr^2$ = semipartial correlation; VC = verbal comprehension; VM = verbal memory; WM = working memory. *$p < .05$. **$p < .01$.

### Table 3. Sequential Regression of Age (Reciprocal), Cognitive Function, and Memory on Level Achieved on IVR Tasks

<table>
<thead>
<tr>
<th>Variables</th>
<th>Level achieved on IVR</th>
<th>Age (reciprocal)</th>
<th>WM</th>
<th>AM</th>
<th>PR</th>
<th>VC</th>
<th>PS</th>
<th>VM</th>
<th>$B$</th>
<th>$SEB$</th>
<th>$\beta$</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>$sr^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.12</td>
<td>.17</td>
<td>.31</td>
<td>.36</td>
<td>.40</td>
<td>.46</td>
<td>.56</td>
<td>.40</td>
<td>.46</td>
<td>-.02</td>
<td>.05</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>WM</td>
<td>.30</td>
<td>.17</td>
<td>.31</td>
<td>.36</td>
<td>.40</td>
<td>.46</td>
<td>.56</td>
<td>.40</td>
<td>.46</td>
<td>-.02</td>
<td>.05</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>AM</td>
<td>.25</td>
<td>.32</td>
<td>.40</td>
<td>.28</td>
<td>.46</td>
<td>.46</td>
<td>.56</td>
<td>.40</td>
<td>.46</td>
<td>-.02</td>
<td>.05</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>PR</td>
<td>.25</td>
<td>.32</td>
<td>.40</td>
<td>.28</td>
<td>.46</td>
<td>.46</td>
<td>.56</td>
<td>.40</td>
<td>.46</td>
<td>-.02</td>
<td>.05</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>VC</td>
<td>.23</td>
<td>.45</td>
<td>.37</td>
<td>.35</td>
<td>.56</td>
<td>.40</td>
<td>.01</td>
<td>.04</td>
<td>.04</td>
<td>-.02</td>
<td>.05</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>PS</td>
<td>.20</td>
<td>.44</td>
<td>.31</td>
<td>.47</td>
<td>.50</td>
<td>.54</td>
<td>.34</td>
<td>.34</td>
<td>.46</td>
<td>-.02</td>
<td>.05</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.47</td>
<td>4.47</td>
<td>.46</td>
<td>.70</td>
<td>.55</td>
<td>.29</td>
<td>.23</td>
<td>.23</td>
<td>.23</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>SD</td>
<td>6.89</td>
<td>.01</td>
<td>40.56</td>
<td>81.28</td>
<td>55.24</td>
<td>77.14</td>
<td>78.14</td>
<td>44.43</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.16</td>
<td>.16</td>
<td>.40</td>
<td>.28</td>
<td>.46</td>
<td>.46</td>
<td>.50</td>
<td>.50</td>
<td>.46</td>
<td>-.02</td>
<td>.05</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
</tr>
</tbody>
</table>

**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. AM = auditory memory; B = unstandardized regression coefficient; $\beta$ = standardized regression coefficient; PR = perceptual reasoning; PS = processing speed; SD = standard deviation; SE B = intercept; $sr^2$ = semipartial correlation; VC = verbal comprehension; VM = verbal memory; WM = working memory. *$p < .05$. **$p < .01$. 

---

INTERACTIVE VOICE RESPONSE SYSTEMS IN OLDER PEOPLE

---

499
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 IVR interactions 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 labor 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 IVR 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.

### Table 4. Sequential Regression of Age (Reciprocal), Cognitive Function, and Memory on Number of Errors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Number of errors</th>
<th>Age (reciprocal)</th>
<th>WM</th>
<th>AM</th>
<th>PR</th>
<th>VC</th>
<th>PS</th>
<th>VM</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>sr$_i^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>$-24$</td>
<td>-30.87</td>
<td>17.66</td>
<td>-14</td>
<td>.06</td>
<td>.05</td>
<td>.06**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WM</td>
<td>$-20$</td>
<td>-0.00</td>
<td>.00</td>
<td>-02</td>
<td>.08</td>
<td>.07</td>
<td>.03*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>$-31$</td>
<td>-0.00</td>
<td>.00</td>
<td>-21</td>
<td>.13</td>
<td>.11</td>
<td>.04**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>$-25$</td>
<td>-0.00</td>
<td>.00</td>
<td>-12</td>
<td>.16</td>
<td>.12</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VC</td>
<td>$-28$</td>
<td>-0.00</td>
<td>.00</td>
<td>-13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>$-24$</td>
<td>-0.00</td>
<td>.00</td>
<td>-02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM</td>
<td>$-20$</td>
<td>-0.00</td>
<td>.00</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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. AM = auditory memory; $B = $unstandardized regression coefficient; $\beta = $standardized regression coefficient; PR = perceptual reasoning; PS = processing speed; SD = standard deviation; SE $B = $intercept; sr$_i^2 = $semipartial correlation; VC = verbal comprehension; VM = verbal memory; WM = working memory.

*p $< .05$. **p $< .01$. 

(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 memory and working memory were 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 memory, and auditory memory) explained approximately the same proportion of the variance.

Taken together, the results of our study indicated that adults aged 65 and older 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 with 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 with 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” was 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 auditory 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, because IVR systems use prerecorded 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 prerecorded messages. This reliance might explain the pre-eminence of auditory memory 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 multifactorial 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 and 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 memory 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 (STPSC/356771-2007) and by TelAsk Technologies.

**Acknowledgments**

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

**Correspondence**

Correspondence should be addressed to Claude Messier, PhD, School of Psychology, University of Ottawa, 136 Jean-Jacques Lussier, Ottawa, Ontario K1N 6N5, Canada. E-mail: cmessier@uottawa.ca.

**References**


