Investigating Voice-Activated Conversation Assistant Interactions with Older Adults

Kayla Chan

Department of Psychology, University of Minnesota

UROP Final Report

Dr. Wilma Koutstaal

September 4, 2021

Aging is associated with declines in physical and cognitive abilities, so technology that aids in scheduling or maintaining autonomy is potentially beneficial for older adults. Voice-activated conversation assistants (CA) are one such technology that could provide greater independence and assist with reminders and planning since they are more accessible to users who may feel embarrassed about their cognitive abilities. While holding great promise, CAs are currently lacking in sensitivity to user preferences and have difficulty with conversational engagement (Tabassum et al., 2019). The project I worked on this term examined characteristics of older adult interactions with CAs using a Wizard of Oz (WOZ) paradigm, as part of a larger Grand Challenge and "EAGER" project conducted by Professor Koutstaal and an interdisciplinary research team that aims to research and develop a CA for older adults, and increase the CA's usability and longer-term adoption by older individuals.

The study used data from a WOZ style study conducted over Zoom with older adult participants. Participants interacted with a CA prototype that, unbeknownst to them, was run by a researcher (i.e., “wizard”) emulating a CA. Due to Covid-19 restrictions on in-person testing, all sessions were conducted on the Zoom audio-video conferencing platform. Participants took part in eight sessions with the CA over four days, answering prospective or retrospective questions about their daily schedule, positive events, and thoughts on various topics. During either Day Two or Day Three, sessions contained a greater number of self-disclosure or positive expression inducing prompts than baseline levels in Days One and Four, with the order of the day consisting of the manipulation randomly determined. For all days, likelihood to use the CA again, and feelings toward the CA (e.g., the CA was pleasant to be with) were recorded at the end of each session and general mood was recorded before and after each session.

Following the study, participants' answers were transcribed and anonymized. Originally, the plan was to examine instances of self-disclosure and positive expression, however due to time constraints (participants cancelling, technology issues, etc.) final transcriptions and cross-checking of the CA interactions are still underway, and Linguistic Inquiry Word Count analysis on the CA interactions will be conducted in the upcoming weeks. So, the focus of this report will be on participants' demographics and feelings towards the CA, as well as the Wizard of Oz paradigm and self-disclosure manipulation.

There were 17 participants (13 female, 4 male), 14 of whom completed all 8 sessions and 3 of whom completed 6 sessions, for a total of 130 sessions. The average age was 69.41 (SD = 5.46), and average years of education was 17.47 years (SD = 1.62). Participants were given questionnaires on technology engagement and adoption, SmartHome system and smartphone use, and competence with technology. In general, a majority of participants reported an interest in engaging with technology, including smartphones, tablets, and computers, as well as proficiency in technology-related tasks such as emailing, searching the web, and social networking. Regarding technology adoption, 15 participants reported using a computer or tablet daily, 14 used a smartphone daily, and 9 used or had previously used a SmartHome device. Broadly, participants were relatively familiar with technology and willing to engage with it.

General mood was recorded before and after each session, measured on a scale of one (not very good at all) to seven (very good). In general, participants responded that they were feeling quite good, grand mean = 6.37, 95% CI [6.10, 6.65]. Averaging across all the sessions that participants completed, there was a modest but significant increase in their mood ratings from beginning to ending interactions, $F(1, 15) = 6.42$, $p = .023$, partial eta squared = .30.

On either Day Two (sessions 3 and 4) or Day Three (sessions 5 and 6), sessions included a greater number of self-disclosure or positive expression inducing prompts (e.g. “tell me a moment in the day that made you smile") than in Days One and Four. This represented the experimental manipulation. There was no main effect of condition, that is, whether the experimental session with additional positive-eliciting prompts from the CA occurred in Sessions 3+4, or in Sessions 5+6, $F = 1.09$ (means = 6.24 and 6.51 respectively). There also was no condition x timepoint interaction, $F = 1.21$, although numerically the increase was slightly more apparent for the Sessions 3+4 condition, who also gave slightly numerically lower ratings overall (begin and end means for Sessions 3+4, = 6.13 and 6.35 respectively; begin and end means for Sessions 5+6 = 6.46 and 6.55, respectively).

Participants also completed a set of 15 questions regarding their feelings towards the CA after each session (see Table 1). All items were on a 5-point scale (1 = completely disagree, 5 = completely agree). Only the 14 participants who completed all 8 sessions (n=9 for Sessions3+4, n=5 for Sessions5+6) were used in the following analyses. In general, participants agreed that interactions with the CA were pleasant and that the CA responded appropriately, with mean responses for all items ranging from 3.47 to 4.73. For the questions 3 ("I felt comfortable sharing personal information during the interaction") and 8 ("I would like to interact with the CA again"), the responses trended toward a linear decline across sessions, question 5 ("I felt involved in the interaction") had a significant effect of session (but the linear effect was only marginal), and question 12 ("The CA communicated properly") had a significant omnibus effect of session. However, for all other questions, there was no significant decline across sessions. In a previous study, the frequency of self-disclosure decreased over time and users found interactions

increasingly less enjoyable, so this result is encouraging and may suggest that the current design is more engaging (Croes & Antheunis, 2020).

At the end of the study, participants were debriefed on the nature of the study and answered a series of questions about their thoughts on the CA (see Table 2). The WOZ design seemed to work well, as thirteen people did not suspect a human was guiding the CA and nine said they were surprised to find a human behind the CA. Some participants noted that they assumed that the technology that would enable a CA such as this to actually exist was already available, and twelve participants agreed that a CA like the one they interacted with could actually be built. When asked which aspects of language would be most difficult to make a Scheduling Assistant handle well, participants gave a range of answers, with several mentioning responding sensitively to emotional or complex responses or idiosyncrasies in the user's speech, such as slang or dialect. Finally, participants were asked to provide feedback on the study as a whole. Participants gave generally positive notes, with several mentioning they could see the CA being useful for older adults.

While the scope of this project did not focus on self-disclosure and positive expression, the project did achieve the objective of testing a CA design and prompts with older adults and gaining valuable feedback about the usability and somewhat longer-term user engagement levels with a CA. Overall, this project has been an invaluable opportunity to learn about and get involved in different aspects of the research process, whether that was transcribing interviews, helping to design questions or phrases for the CA, or reading papers that gave background to the study. This bigger picture of research, from nitty-gritty details to broader theories or reasonings for making certain choices, has also been a learning experience for me. Through this project, I

think I've been able to understand that research needs to be careful and well thought-out at every level and to apply that to my own studies and projects as well.

Another important takeaway from the project is the importance of communicating and working well with a team. The research team was led by Dr. Koutstaal, but each member had a role to play; collaborating together and having different perspectives made the project richer. In sum, this project has been both a valuable academic experience and an opportunity to learn and grow. I'm excited to see where the rest of the Grand Challenge and EAGER project goes and am hopeful that the work we've done so far and the eventual CA prototype will be helpful for the older adult population.

Table 1
Post-Session Questionnaire

Question Number	Mean	95% Confidence Interval	Question
1	4.69	4.40, 4.99	The CA was pleasant to be with
2	4.73	4.45, 5.01	The CA was sociable with me
3	4.47	3.94, 5.01	I felt comfortable sharing personal information during the interaction
4	4.48	4.03, 4.93	I felt like I could be open during the interaction
5	4.49	4.02, 4.96	I felt involved in the interaction
6	4.27	3.64, 4.90	I enjoyed the interaction with the CA
7	4.24	3.79, 4.68	I considered the interaction with the CA to be smooth
8	3.99	3.23, 4.75	I would like to interact with the CA again
9	4.11	3.51, 4.71	I feel the interaction with the CA was satisfying
10	4.17	3.58, 4.76	The CA said the right thing to make me feel better
11	4.45	3.95, 4.94	The CA responded appropriately to my feelings and emotions
12	4.56	4.13, 5.00	The CA communicated properly
13	4.63	4.28, 4.98	The CA came across as competent
14	4.11	3.57, 4.66	The CA came across as natural
15	3.47	2.72, 4.21	Overall, I would like to have a longer conversation with the CA

Table 2
Debrief Questionnaire

Response Scale	Response	Question
(1 = Not at all, 4 = Neutral, 7 = Very much)	very surprised (6 or 7): 9 intermediate (3, 4, or 5): 3 not surprised (1 or 2): 5	How surprised are you to learn that there was a human respondent who was directing the Scheduling Assistant?
(1 = Not at all, 4 = Neutral, 7 = Very much)	very much so (6 or 7): 12 intermediate (3, 4, or 5): 5	Do you think that a system as responsive as the one you experienced during this last session could actually be built?
Yes or No	No: 13 participants Yes: 4 participants	At any time in all the sessions, did you suspect that a human respondent might be guiding the Scheduling Assistant?
Free response	Various	Did any specific aspects of your interactions with the Scheduling Assistant seem to you to be more similar to interacting with a human respondent than to interacting with a computer-based system?
Free response	Various	Which aspects of language, if any, do you think it would be most difficult to make a Scheduling Assistant handle well?
Free response	Various	Please provide any comments or feedback you would like the research team to consider regarding any aspects of the experiment, including the use of the Wizard of Oz format.

References

Croes, E. A. J, & Antheunis, M. L. (2020). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*. https://doi.org/10.1177/0265407520959463

Tabassum, M., Kosiński, T., Frik, A., Malkin, N., Wijesekera, P., Egelman, S., & Lipford, H. (2019). Investigating users' preferences and expectations for always-listening voice assistants. *Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3*(4), 1-23. https://doi.org/10.1145/3369807