Usable and Accessible Voice Technology for People with Vision Loss

Hyung Nam Kim and Boubacar Oumarou
Department of Industrial and Systems Engineering, North Carolina A&T State University,
Greensboro, NC 27411, USA

INTRODUCTION

Today, the popularity of voice-enabled technology emerges and is expected to continue to grow. Yet, there is a lack of understanding on user interface and interaction designs of the voice technology suitable for people with vision loss, especially older adults. As the underlying mechanisms are unclear about the combined effects of vision loss and aging on memory span of auditory information, researchers and professionals are likely to encounter a challenge with developing a user-centered design of voice technology suitable for older adults with vision loss. This study aims to investigate the memory span of older adults with vision loss and to explore user-friendly voice technology designs for them.

METHODS

This study invited a convenience sample of 20 older adults (ages 65 and older) with vision loss (visual acuity worse than 20/70). Word memory span testing was conducted auditorily to examine the degree to which participants correctly recall words. We investigated user preferences for voice user interfaces and interactions via the Wizard-of-Oz method.

RESULTS

The word memory span was 3 on average. The majority of the participants were able to recall up to 3 words. A few participants could recall up to 4 words, i.e., two participants recalled a single set of 4 words while one participant recalled two sets of 4 words. The majority of the participants preferred the combined interaction of *inbound and outbound calls*, followed by the interaction of *inbound call only*. However, none of the participants appreciated the interaction of *outbound call only*. Many of them would then like to interact with the system via the *interactive multi-level menu* mode, followed by the *listen-all-menu* mode and the *yes/no simple menu* mode. Nearly half of the participants preferred to be informed of the data categorized by *activities of daily living*, followed *by timeframe* and *listening to everything without any category*. The duration of vision loss of the participants who preferred structured information architecture tend to be longer than that of their peers who preferred unstructured one.

DISCUSSION

The word memory span of the participants was short as it was 3. Thus, it is recommended that the number of voice navigation structures be limited to three layers. In terms of user interaction, the participants appreciated user empowerment of controlling the system. For instance, they preferred the inbound call system in that they would not like to simply wait until they receive a call from the system. The participants also appreciated the system that could categorize the data *by daily living activities* as they would like to avoid wasting their time in listening to all contents. Yet, another group appreciated the system that could categorize the data *by time*. They also suggested a hybrid format; that is, the system asks users to choose both a particular time frame and data type, which would contribute to narrowing down the list of data that users have to listen to. In contrast to the aforementioned groups, the other group stated that the option of listening to everything is consistent with their mental model and decision-making process; that is, they typically hear everything to make a decision.

Another finding is that the participants with longer duration of vision loss preferred a complex, structured information architecture as compared to their peers with shorter duration. It may be argued that such a long period of time would be sufficient for those with vision loss to augment their ability of information processing and performing complex tasks (e.g., comprehension and interaction with multiple navigation menus).

This study was undertaken to advance knowledge of the memory span of older adults with vision loss and user preferences for voice technology. The results of this study lead to valuable insights into user-friendly voice user interface and interaction designs that are likely applicable to many other voice technology applications.

Acknowledgements: This material is based upon work supported by the National Science Foundation under Grant No. 1831969.