Application of Speech Technology in a Home Based Assessment Kiosk for Early Detection of Alzheimer's Disease

Rachel Coulston, Esther Klabbers, Jacques de Villiers and John-Paul Hosom

Center for Spoken Language Understanding
OGI School of Science & Engineering at OHSU
20000 NW Walker Road, Beaverton, OR 97006, USA
rachel@cslu.ogi.edu

Abstract

Alzheimer's disease, a degenerative disease that affects an estimated 4.5 million people in the U.S., can be treated far more effectively when it is detected early. There are numerous challenges to early detection. One is objectivity, since caretakers are often emotionally invested in the health of the patients, who may be their family members. Consistency of administration can also be an issue, especially where longitudinal results from different examiners are compared. Finally, the frequency of testing can be adversely affected by scheduling or cost constraints for in-home psychometrician visits. The kiosk system described in this paper, currently deployed in homes around the country, uses speech technology to provide advantages that address these challenges.

Index Terms: spoken language systems, in-home psychometric testing, Alzheimer's disease.

1. Introduction

Alzheimer's Disease (AD) is a widespread and growing concern among an increasingly larger aging population. It affects an estimated 4.5 million people in the U.S. It increases dramatically with age, affecting approximately 40-50% of people age 85 and older. There is no cure, but progression of the disease can be slowed by therapies that are currently available, with increased success if intervention is early. Early detection of a gradual, degenerative disease such as this, however, is difficult. Most individuals' declines only begin to be noticeable after devastating brain loss has already occurred. Challenges central to this problem are: frequency of monitoring, objectivity of caregivers, and subtlety of the earliest signs.

Seniors are at highest risk for developing AD, but are also the most likely to have problems with mobility, disorientation, and other health complications. In-home testing is therefore the best option, but regularly sending a psychometrician to the home of each patient for preventative monitoring can be extremely costly. Consistency of administration across instances is another issue, especially given the longitudinal nature of early detection monitoring of AD. Caregivers have the most consistent contact with patients, but may lack objectivity or adapt to changes before noticing them and reporting them to physicians. A computerized in-home monitoring system addresses these issues [1].

To acheive the goals of a longitudinal AD study, subjects must continue to participate. Drop-out rates are estimated at 40 percent over the course of a typical 4-year study [2]. There are several reasons the kiosk system is expected to improve upon this projected subject loss. The physical presence of the kiosk in the home not only serves as a reminder, but makes study compliance quite easy for the subject, once he is enrolled. Reminders and data collection sessions themselves are initiated by the kiosk system, and do not rely on the subject to complete items at his leisure. These facts are expected to improve subject retention. Even small improvements in retention would be valuable, especially for studies that aim to track subjects over very long periods of time.

Figure 1: *Kiosk and user during a session.*

One of the practical end goals of a kiosk system such as the one we describe in this paper, is to facilitate the collection of efficacy data for AD drug treatment trials, since the administration of these tests as conducted on the computer is standardized across all participants and visits, unlike in the conventional in-person style of testing. Analysis is also completed taking full advantage of objectivity, something which can be difficult for people scoring in situations where they have contact with subjects.

2. The home based assessment study

The kiosk system described in this paper is part of a larger study in which three methods of testing are being compared: conventional, telephone and computer testing. Conventional testing consists of an in-person administration of the psychometric battery, paired with mail-in questionnaires. Research study coordinators can call to remind the participants to mail in their questionnaires. Telephone-based testing is done via an interactive voice response (IVR) system [3]. Participants are provided with a large-button, hearing-aid compatible telephone. An appointment is made by the study coordinator for the IVR system to initiate a phone call to the participant at a specific date and time, and questionnaires and psychometric tests are administered and recorded via telephone. The computerized test using a kiosk will be further described in this paper.

Currently, a pilot study is underway to compare the three testing methods. For the pilot study, a total of 45 participants have been recruited who are randomly assigned to participate in the conventional, IVR, or kiosk test (15 each). At conference time, the main study will be in progress, with data being collected in 600 homes, of which 200 homes will be outfitted with kiosks such as the one we describe here.

3. The kiosk

The kiosk system is designed to be used in the home, with tests completed monthly. A site visit by the research study coordinator is required to set up the kiosk, schedule appointment times, and walk the user through his initial training. A short video will show the participant an example of the test they are to take. The actual testing occurs unattended in the participant's own home, which decreases the number of costly and intrusive site visits needed to monitor memory function and also aims to increase the amount and quality of data gathered.

3.1. Physical requirements

The physical footprint of the kiosk is small, so it will not take up much space in the participant's home. The participant receives only a compact PC and a flat panel monitor with a telephone receiver on one side. There is no dedicated keyboard or mouse attached to the system. The study coordinator will use both briefly for setup, and remove them from the premises when she leaves. The subject interacts with the system solely via speech through the handset or via touchscreen manipulation. All hardware for the kiosk is commercially available.

- Computer:
  - Dell OptiPlex 745 Ultra Small Form Factor
  - Celeron 346/3.06GHz processor
  - 512MB memory
- Monitor:
  - Planar PT1700MU 17-inch USB Touchscreen LCD
- Audio Advantage Micro USB Audio Adapter
- Handset (requires modification): TeleVoIP PC Handset
- Internet connection

All components are used as they arrive off-the-shelf with one exception. We alter the handset wiring to enable the system to be aware of the state of the hook switch, which is free when the user is holding the handset to his ear, and depressed when the handset is resting in its cradle (or set down on a flat surface, such as a countertop).

3.2. Software

The client software runs under Windows XP, enabling the kiosk to be installed on readily commercially available systems. As the kiosk system computer is not intended to be used for any

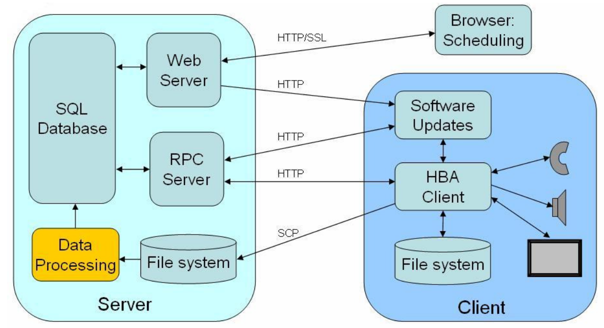

Figure 2: *The overall architecture of the kiosk data collection system.*

other purpose, such as emailing, web browsing, etc., the software runs automatically and continuously and prevents the machine from being used for other activities. In the event of a power failure, it will start up automatically upon reboot, and can be monitored remotely by study personnel. Figure 2 shows a diagram of the software architecture. Data is cached on the local file system on the client machine, and uploaded to the server at regular intervals. Appointments are scheduled either in person or by phone, and the selected times are entered by the study coordinators via a web interface. The kiosk contacts the server every 5 minutes to check for newly scheduled testing appointments. In the event that the network connection becomes unavailable, any tests the local machine has been informed of already will proceed without interruption or data loss. The test data will be stored locally until the internet service is restored, at which point it will resume its normal checking intervals, software updates, and it will upload the data it has stored. On the server, data is stored in a MySQL database running on Red Hat Enterprise Linux.

3.3. Interface

The kiosk system rings and displays a visual reminder of the appointment time 72 hours prior to a scheduled test. Two days before the test, a silent message is displayed on the screen for one hour. Twenty-four hours before the test a (silent) text message is displayed for 5 minutes every hour, on the hour between 07:00 and 20:00 local time. The kiosk monitor is turned off the rest of the time. A session begins with the kiosk ringing at the scheduled time. Hook switch state and speech detection are used to determine when the participant has answered, at which point the user begins to be guided through his session by the video experimenter. Each instruction to the user is played as a video clip of the experimenter, so the subject sees and hears just what would be presented during an in-person visit. The audio is played either via the handset (if lifted) or the USB audio adapter that is attached to the computer (if handset is resting in cradle). Wherever the tests allow it, summary instructions are also shown in text at the top of the screen. This redundancy eases the burden on hearing- and vision-impaired users, by providing context in other modalities with which to interpret the limited amount they can hear or see.

The use of a finger-sensitive touchscreen obviates the need for a stylus, which can be lost or misplaced (and sometimes complicated to extract from its storage compartment). Using a fingertip to point or trace is also a highly natural action, even for those who have not been exposed to touchscreen technology

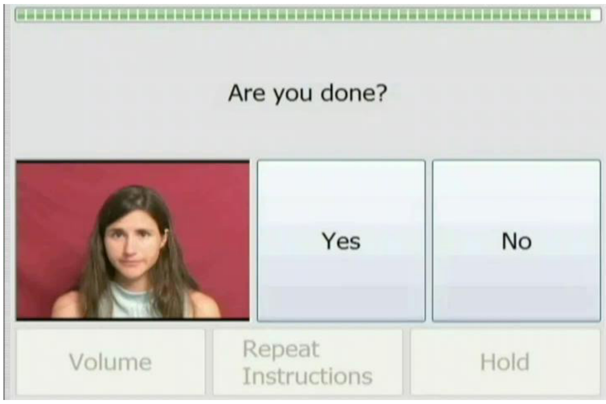

Figure 3: *One possible state of the interactive touchscreen, showing the video experimenter.*

before. Additionally, speech recognition is enabled for simple navigation through the interface. For example, the yes/no question in Fig. 3 can be answered either by pressing one of the buttons on the touchscreen or by speaking the word yes or no aloud into the telephone-style handset.

4. Overview of the test battery

For all three methods of testing, the test includes a diverse battery of questionnaires and cognitive tests. With the kiosk, an average test session takes approximately 30 minutes of the participant's time to complete. In this paper, we limit our discussion to those tasks that currently use speech technology and those for which speech technology is under development. Questionnaires and tasks not discussed here include self-report questions, also delivered by the kiosk, about the quality of life, medication adherence, how well participants are able to complete activities of daily life and so on. Another nonspeech task in the cognitive battery requires participants to connect labeled dots in the correct sequence using their finger on the touchscreen. In the following sections, four tasks using speech technology will be discussed in more detail.

4.1. Word list recall

For the word list recall task, ten words are presented visually and the subject reads each one aloud as it appears. At the end of the presentation, the subject is asked to recall as many of the words as he can (in any order). The same stimuli are then presented two more times. The presentation order is varied for each repeat. After a delay during which separate tasks are completed, the subject is again asked to recall the words (although this time they are not presented immediately before the recall is requested).

Speech detection is used to determine whether the subject is following the instruction to read each word aloud. If there is silence, the experimenter reminds the subject to read each word aloud. During recall, speech and silence detection are used to determine when to prompt subjects to try to think of more words, and ultimately to ask if they are finished.

Speech recognition is being implemented to automatically score this closed class (limited set) of words. Since the subject is expected to say one of ten words, which are known in advance and do not change, the recognizer can be well trained to identify these words. Currently, the subjects' responses are recorded and uploaded to a server for manual scoring, just as would occur during an in-person assessment. Collecting these data in the pilot study means that the manual scores will be available as the "gold standard" for developing and testing the speech recognition software.

4.2. Backward digit span

For the backward digit span test, strings of digits are said aloud by the video experimenter. Subjects are asked to repeat each string in reverse order. For instance, if the stimulus is "2-8-3" the correct response would be "3-8-2". The task begins with strings two digits in length and continues until the subject makes an error on two consecutive strings of the same difficulty level (length) or until the last string is reached, which is eight digits in length.

Participants' responses are recognized and scored on-the-fly. Because digit recognition is very accurate, the flow of this task is controlled by the speech recognition software [4], [5]. Automatic scoring is used to determine whether to discontinue the task or to allow it to become more difficult. Audio of the subject's responses is also recorded, to be used for manual verification of automatic recognition accuracy.

4.3. Category fluency

For the category fluency task, subjects are asked to name as many members of a set as they can think of (i.e., animals) in one minute. Currently, responses to this task are audio-recorded and scored manually. The interface uses speech detection in this task as well, to determine when the experimenter should intervene: either by prompting the participant to try and think of some more members of the set or, eventually, to ask if he is done.

Because this is a fairly limited domain, some scoring (or at least preprocessing) will be possible. Pilot data will be used to establish a dictionary with coverage of the animals which are most commonly named during the task. The subsequent iteration of the kiosk will be able to produce a transcript and a preliminary score. The manual scorer will be able to check the recognizer's word spotting accuracy and make additions or corrections if needed.

Data from this task will also be used to investigate potential features for enhanced detection of impairment. For instance, the data can be analyzed for clusters of related animals, and temporal characteristics of the generated lists, such as latency within clusters versus between clusters. This will be a novel use of data that are already being collected for the cognitive test itself.

4.4. East boston story recall

For the story recall test, a story is told by the video experimenter to the subject, who is subsequently asked to retell it at two different points during the session. The story contains 12 elements that have to be remembered to obtain a perfect score. The first retelling occurs immediately following the presentation and the second retelling happens twenty minutes later, after several intervening tasks have taken place. The participant is instructed to use the same words that were used in the original story to their best ability, and to tell as much of the story as they can remember.

"Three children were alone at home, and the house caught on fire. A brave fireman managed to climb in a back window and carry them to safety. Aside from minor cuts and bruises, all were well."

Audio is recorded as the participant responds and speech detection determines when the experimenter should intervene, prompting the subject to try to recall more, then asking if he is done. Because the story is fixed, recognition will be used to spot the words that are supposed to be retold. But since participants may use their own words, automated recognition based scoring will require supervision. In subsequent iterations of the kiosk system, a manual scorer will be able to check the recognizer's word spotting results and make additions or corrections if needed.

In addition, future work is planned, not just to enhance automatic scoring, but also expand the investigation of speech markers of AD. This will involve leveraging the recording to extract more information about a subject's performance than current scoring techniques. An example of this would be weighting the 12 different story elements differently in determining the overall recall quality, rather than just counting the number of elements they recalled. Retellings of stories can be further analyzed for patterns of retelling, such as temporal ordering of events, or particular story elements that are typically omitted from the retellings by particular groups.

5. Conclusions

The kiosk is designed to be used as a prevention instrument for early detection of Alzheimer's Disease. It can facilitate collection of data for high-risk elderly populations, which can aid diagnosis of early cognitive decline or onset of dementia. The combination of speech processing, audio recording for subsequent human scoring, and multi-modal redundancy in stimulus presentation all help to streamline data collection and scoring. All hardware is commercially available, and minor modifications are needed to the handset, enabling this software to be used for large data collections. The invariability in the delivery of instrutions and stimuli are valuable for keeping stimulus presentation consistent over the course of longitudinal studies. Additionally, unlike human psychometricians, the kiosk will not be subject to the influence of any prejudice about whether or not cognitive decline is present (e.g., assumptions about cognitive function that might be made unconsciously based upon physical appearance of health). Thus, the kiosk system that we present in this paper meets the consistency and objectivity demands we discussed. It should become a convenient, accurate, cost-saving system for early diagnosis and longitudinal monitoring of Alzheimer's Disease.

6. Acknowledgements

We are grateful to the funders of the Alzheimer's Disease Cooperative Study (ADCS), the National Institutes of Health's (NIH) National Institute on Aging (NIA), as well as Intel's Behavioral Assessment and Intervention Commons (BAIC) program for providing additional funding. The views expressed in this paper do not necessarily represent the views of the NIH or of Intel. We would also like to acknowledge Tamara Hayes and Jeff Kaye and their teams for their work on this project. Jessica Payne-Murphy's generous contribution of her voice and likeness for use as our video experimenter is particularly appreciated.

7. References

[1] T.L. Hayes, M. Pavel, J.A. Kaye, (2004),"An unobtrusive in-home monitoring system for detection of key motor changes preceding cognitive decline", In *Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA*.

[2] J.C. Mundt, & M. King, (2003). "Predicting early treatment drop out using interactive voice response (IVR)" *Alcoholism: Clinical and Experimental Research*,27, 28A.

[3] J.C. Mundt, K.L. Ferber, M. Rizzo, J.H. Greist, (2001), "Computer-automated dementia screening using a touchtone telephone" *Arch Intern Med.*, 161:2481-2487.

[4] P. Cosi, J.P. Hosom, J. Schalkwyk, S. Sutton, and R.A. Cole, (1998), "Connected digit recognition experiments with the OGI Toolkit's neural network and HMM-based recognizers", In *Proceedings of IVTTA'98. 4th IEEE Workshop on Interactive Voice Technology for Telecommunications Applications, Torino, Italy*, p 135-140.

[5] J.P. Hosom, (2000), "Automatic Time Alignment of Phonemes Using Acoustic-Phonetic Information", *Ph.D. thesis, Oregon Graduate Institute of Science and Technology (now OGI School of Science & Engineering at Oregon Health & Science University).*