Speech recognition in Alzheimer's disease with personal assistive robots

Frank Rudzicz[1,2,*] and **Rosalie Wang**[1] and **Momotaz Begum**[3] and **Alex Mihailidis**[2,1]

[1] Toronto Rehabilitation Institute, Toronto ON; [2] University of Toronto, Toronto ON;
[3] University of Massachussetts Lowell
*frank@cs.toronto.edu

Abstract

To help individuals with Alzheimer's disease live at home for longer, we are developing a mobile robotic platform, called ED, intended to be used as a personal caregiver to help with the performance of activities of daily living. In a series of experiments, we study speech-based interactions between each of 10 older adults with Alzheimers disease and ED as the former makes tea in a simulated home environment. Analysis reveals that speech recognition remains a challenge for this recording environment, with word-level accuracies between 5.8% and 19.2% during household tasks with individuals with Alzheimer's disease. This work provides a baseline assessment for the types of technical and communicative challenges that will need to be overcome in human-robot interaction for this population.

1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder primarily impairing memory, followed by declines in language, ability to carry out motor tasks, object recognition, and executive functioning (American Psychiatric Association, 2000; Gauthier et al., 1997). An accurate measure of functional decline comes from performance in activities of daily living (ADLs), such as shopping, finances, housework, and self-care tasks. The deterioration in language comprehension and/or production resulting from specific brain damage, also known as aphasia, is a common feature of AD and other related conditions. Language changes observed clinically in older adults with dementia include increasing word-finding difficulties, loss of ability to verbally express information in detail, increasing use of generic references (e.g., "it"), and progressing difficulties understanding information presented verbally (American Psychiatric Association, 2000).

Many nations are facing healthcare crises in the lack of capacity to support rapidly aging populations nor the chronic conditions associated with aging, including dementia. The current healthcare model of removing older adults from their homes and placing them into long-term care facilities is neither financially sustainable in this scenario (Bharucha et al., 2009), nor is it desirable. Our team has been developing "smart home" systems at the Toronto Rehabilitation Institute (TRI, part of the University Health Network) to help older adults "age-in-place" by providing different types of support, such as step-by-step prompts for daily tasks (Mihailidis et al., 2008), responses to emergency situations (Lee and Mihaildis, 2005), and means to communicate with family and friends. These systems are being evaluated within a completely functional re-creation of a one-bedroom apartment located within The TRI hospital, called HomeLab. These smart home technologies use advanced sensing techniques and machine learning to autonomously react to their users, but they are fixed and embedded into the environment, e.g., as cameras in the ceiling. Fixing the location of these technologies carries a tradeoff between utility and feasibility – installing multiple hardware units at all locations where assistance could be required (e.g., bathroom, kitchen, and bedroom) can be expensive and cumbersome, but installing too few units will present gaps where a user's activity will not be detected. Alternatively, integrating personal mobile robots with smart homes can overcome some of these tradeoffs. Moreover, assistance provided via a physically embodied robot is often more acceptable than that provided by an embedded system (Klemmer et al., 2006).

With these potential advantages in mind, we conducted a 'Wizard-of-Oz' study to explore the

feasibility and usability of a mobile assistive robot that uses the step-by-step prompting approaches for daily activities originally applied to our smart home research (Mihailidis et al., 2008). We conducted the study with older adults with mild or moderate AD and the tasks of hand washing and tea making. Our preliminary data analysis showed that the participants reacted well to the robot itself and the prompts that it provided, suggesting the feasibility of using personal robots for this application (Begum et al., 2013). One important identified issue is the need for an automatic speech recognition system to detect and understand utterances specifically from older adults with AD. The development of such a system will enable the assistive robot to better understand the behaviours and needs of these users for effective interactions and will further enhance environmental-based smart home systems.

This paper presents an analysis of the speech data collected from our participants with AD when interacting with the robot. In a series of experiments, we measure the performance of modern speech recognition with this population and with their younger caregivers with and without signal preprocessing. This work will serve as the basis for further studies by identifying some of the development needs of a speech-based interface for robotic caregivers for older adults with AD.

2 Related Work

Research in smart home systems, assistive robots, and integrated robot/smart home systems for older adults with cognitive impairments has often focused on assistance with activities of daily living (i.e., reminders to do specific activities according to a schedule or prompts to perform activity steps), cognitive and social stimulation and emergency response systems. Archipel (Serna et al., 2007) recognizes the user's intended plan and provides prompts, e.g. with cooking tasks. Autominder, (Pollack, 2006), provides context-appropriate reminders for activity schedules, and the COACH (Cognitive Orthosis for Assisting with aCtivities in the Home) system prompts for the task of hand-washing (Mihailidis et al., 2008) and tea-making (Olivier et al., 2009). Mynatt et al. (2004) have been developing technologies to support aging-in-place such as the Cooks Collage, which uses a series of photos to remind the user what the last step completed was if the user is interrupted during a cooking task. These interventions tend to be embedded in existing environments (e.g., around the sink area).

More recent innovations have examined integrated robot-smart home systems where systems are embedded into existing environments that communicate with mobile assistive robots (e.g., CompanionAble, (Mouad et al., 2010); Mobiserv Kompai, (Lucet, 2012); and ROBADOM (Tapus and Chetouani, 2010)). Many of these projects are targeted towards older adults with cognitive impairment, and not specifically those with significant cognitive impairment. One of these systems, CompanionAble, with a fully autonomous assistive robot, has recently been tested in a simulated home environment for two days each with four older adults with dementia (AD or Pick's disease/frontal lobe dementia) and two with mild cognitive impairment. The system provides assistance with various activities, including appointment reminders for activities input by users or caregivers, video calls, and cognitive exercises. Participants reported an overall acceptance of the system and several upgrades were reported, including a speech recognition system that had to be deactivated by the second day due to poor performance.

One critical component for the successful use of these technological interventions is the usability of the communication interface for the targeted users, in this case older adults with Alzheimer's disease. As in communication between two people, communication between the older adult and the robot may include natural, freeform speech (as opposed to simple spoken keyword interaction) and non-verbal cues (e.g., hand gestures, head pose, eye gaze, facial feature cues), although speech tends to be far more effective (Green et al., 2008; Goodrich and Schultz, 2007). Previous research indicates that automated communication systems are more effective if they take into account the affective and mental states of the user (Saini et al., 2005). Indeed, speech appears to be the most powerful mode of communication for an assistive robot to communicate with its users (Tapus and Chetouani, 2010; Lucet, 2012).

2.1 Language use in dementia and Alzheimer's disease

In order to design a speech interface for individuals with dementia, and AD in particular, it is

important to understand how their speech differs from that of the general population. This then can be integrated into future automatic speech recognition systems. Guinn and Habash (2012) showed, through an analysis of conversational dialogs, that repetition, incomplete words, and paraphrasing were significant indicators of Alzheimer's disease relative but several expected measures such as filler phrases, syllables per minute, and pronoun rate were not. Indeed, pauses, fillers, formulaic speech, restarts, and speech disfluencies are all hallmarks of speech in individuals with Alzheimer's (Davis and Maclagan, 2009; Snover et al., 2004). Effects of Alzheimer's disease on syntax remains controversial, with some evidence that deficits in syntax or of agrammatism could be due to memory deficits in the disease (Reilly et al., 2011).

Other studies has applied similar analyses to related clinical groups. Pakhomov et al. (2010) identified several different features from the audio and corresponding transcripts of 38 patients with frontotemporal lobar degeneration (FTLD). They found that pause-to-word ratio and pronoun-to-noun ratios were especially discriminative of FTLD variants and that length, hesitancy, and agramatism correspond to the phenomenology of FTLD. Roark et al. (2011) tested the ability of an automated classifier to distinguish patients with mild cognitive impairment from healthy controls that include acoustic features such as pause frequency and duration.

2.2 Human-robot interaction

Receiving assistance from an entity with a physical body (such as a robot) is often psychologically more acceptable than receiving assistance from an entity without a physical body (such as an embedded system) (Klemmer et al., 2006). Physical embodiment also opens up the possibility of having more meaningful interaction between the older adult and the robot, as discussed in Section 5.

Social collaboration between humans and robots often depends on communication in which each participant's intention and goals are clear (Freedy et al., 2007; Bauer et al., 2008; Green et al., 2008). It is important that the human participant is able to construct a useable 'mental model' of the robot through bidirectional communication (Burke and Murphy, 1999) which can include both natural speech and non-verbal cues

(e.g., hand gestures, gaze, facial cues), although speech tends to be far more effective (Green et al., 2008; Goodrich and Schultz, 2007).

Automated communicative systems that are more sensitive to the emotive and the mental states of their users are often more successful than more neutral conversational agents (Saini et al., 2005). In order to be useful in practice, these communicative systems need to mimic some of the techniques employed by caregivers of individuals with AD. Often, these caregivers are employed by local clinics or medical institutions and are trained by those institutions in ideal verbal *communication strategies* for use with those having dementia (Hopper, 2001; Goldfarb and Pietro, 2004). These include (Wilson et al., 2012) but are not limited to relatively slow rate of speech, verbatim repetition of misunderstood prompts, closed-ended (e.g., 'yes/no') questions, and reduced syntactic complexity. However, Tomoeda et al. (1990) showed that rates of speech that are too slow may interfere with comprehension if they introduce problems of short-term retention of working memory. Small et al. (1997) showed that paraphrased repetition is just as effective as verbatim repetition (indeed, syntactic variation of common semantics may assist comprehension). Furthermore, Rochon et al. (2000) suggested that the syntactic complexity of utterances is not necessarily the only predictor of comprehension in individuals with AD; rather, correct comprehension of the semantics of sentences is inversely related to the increasing number of propositions used – it is preferable to have as few clauses or core ideas as possible, i.e., one-at-a-time.

3 Data collection

The data in this paper come from a study to examine the feasibility and usability of a personal assistive robot to assist older adults with AD in the completion of daily activities (Begum et al., 2013). Ten older adults diagnosed with AD, aged ≥ 55, and their caregivers were recruited from a local memory clinic in Toronto, Canada. Ethics approval was received from the Toronto Rehabilitation Institute and the University of Toronto. Inclusion criteria included fluency in English, normal hearing, and difficulty completing common sequences of steps, according to their caregivers. Caregivers had to be a family or privately-hired caregiver who provides regular

care (e.g., 7 hours/week) to the older adult participant. Following informed consent, the older adult participants were screened using the Mini Mental State Exam (MMSE) (Folstein et al., 2001) to ascertain their general level of cognitive impairment. Table 1 summarizes relevant demographics.

	Sex	Age (years)	MMSE (/30)
OA1	F	76	9
OA2	M	86	24
OA3	M	88	25
OA4	F	77	25
OA5	F	59	18
OA6	M	63	23
OA7	F	77	25
OA8	F	83	19
OA9	F	84	25
OA10	M	85	15

Table 1: Demographics of older adults (OA).
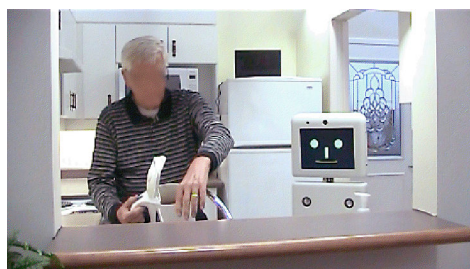
(a)

(b)

Figure 1: ED and two participants with AD during the tea-making task in the kitchen of HomeLab at TRI.

3.1 ED, the personal caregiver robot

The robot was built on an iRobot base (operating speed: 28 cm/second) and both its internal construction and external enclosure were designed and built at TRI. It is 102 cm in height and has separate body and head components; the latter is primarily a LCD monitor that shows audiovisual prompts or displays a simple 'smiley face' other-

wise, as shown in Figure 2. The robot has two speakers embedded in its 'chest', two video cameras (one in the head and one near the floor, for navigation), and a microphone. For this study, the built-in microphones were not used in favor of environmental Kinect microphones, discussed below. This was done to account for situations when the robot and human participant were not in the same room simultaneously.

The robot was tele-operated throughout the task. The tele-operator continuously monitored the task progress and the overall affective state of the participants in a video stream sent by the robot and triggered social conversation, asked task-related questions, and delivered prompts to guide the participants towards successful completion of the tea-making task (Fig. 1).
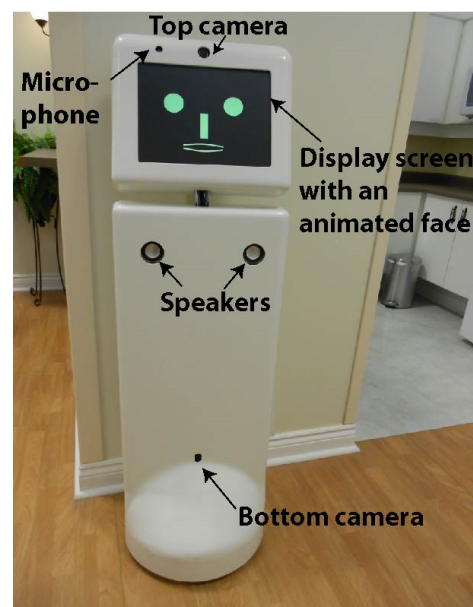

Figure 2: The prototype robotic caregiver, ED.

The robot used the Cepstral commercial text-to-speech (TTS) system using the U.S. English voice 'David' and its default parameters. This system is based on the Festival text-to-speech platform in many respects, including its use of linguistic preprocessing (e.g., part-of-speech tagging) and certain heuristics (e.g., letter-to-sound rules). Spoken prompts consisted of simple sentences, sometimes accompanied by short video demonstrations designed to be easy to follow by people with a cognitive impairment.

For efficient prompting, the tea-making task was broken down into different steps or sub-task. Audio or audio-video prompts corresponding to

each of these sub-tasks were recorded prior to data collection. The human-robot interaction proceeded according to the following script when collaborating with the participants:

1. Allow the participant to initiate steps in each sub-task, if they wish.

2. If a participant asks for directions, deliver the appropriate prompt.

3. If a participant requests to perform the sub-task in their own manner, agree if this does not involve skipping an essential step.

4. If a participant asks about the location of an item specific to the task, provide a full-body gesture by physically orienting the robot towards the sought item.

5. During water boiling, ask the participant to put sugar or milk or tea bag in the cup. Time permitting, engage in a social conversation, e.g., about the weather.

6. When no prerecorded prompt sufficiently answers a participant question, respond with the correct answer (or *"I don't know"*) through the TTS engine.

3.2 Study set-up and procedures

Consent included recording video, audio, and depth images with the Microsoft Kinect sensor in HomeLab for all interviews and interactions with ED. Following informed consent, older adults and their caregivers were interviewed to acquire background information regarding their daily activities, the set-up of their home environment, and the types of assistance that the caregiver typically provided for the older adult.

Participants were asked to observe ED moving in HomeLab and older adult participants were asked to have a brief conversation with ED to become oriented with the robot's movement and speech characteristics. The older adults were then asked to complete the hand-washing and tea-making tasks in the bathroom and kitchen, respectively, with ED guiding them to the locations and providing specific step-by-step prompts, as necessary. The tele-operator observed the progress of the task, and delivered the pre-recorded prompts corresponding to the task step to guide the older adult to complete each task. The TTS system was used to respond to task-related questions and to engage in social conversation. The caregivers

were asked to observe the two tasks and to intervene only if necessary (e.g., if the older adult showed signs of distress or discomfort). The older adult and caregiver participants were then interviewed separately to gain their feedback on the feasibility of using such a robot for assistance with daily activities and usability of the system. Each study session lasted approximately 2.5 hours including consent, introduction to the robot, tea-making interaction with the robot, and post-interaction interviews. The average duration for the tea-making task alone was 12 minutes.

4 Experiments and analysis

Automatic speech recognition given these data is complicated by several factors, including a preponderance of utterances in which human caregivers speak concurrently with the participants, as well as inordinately challenging levels of noise. The estimated signal-to-noise ratio (SNR) across utterances range from -3.42 dB to 8.14 dB, which is extremely low compared to typical SNR of 40 dB in clean speech. One cause of this low SNR is that microphones are placed in the environment, rather than on the robot (so the distance to the microphone is variable, but relatively large) and that the participant often has their back turned to the microphone, as shown in figure 1.

As in previous work (Rudzicz et al., 2012), we enhance speech signals with the log-spectral amplitude estimator (LSAE) which minimizes the mean squared error of the log spectra given a model for the source speech $X_k = A_k e^{(j\omega_k)}$, where A_k is the spectral amplitude. The LSAE method is a modification of the short-time spectral amplitude estimator that finds an estimate of the spectral amplitude, $\hat{A}_k$, that minimizes the distortion

$$E\left[\left(log A_k - \log \hat{A}_k\right)^2\right], \qquad (1)$$

such that the log-spectral amplitude estimate is

$$\hat{A}_k = \exp\left(E\left[\ln A_k \mid Y_k\right]\right)$$
$$= \frac{\xi_k}{1+\xi_k} \exp\left(\frac{1}{2}\int_{v_k}^{\infty} \frac{e^{-t}}{t} dt\right) R_k, \qquad (2)$$

where ξ_k is the *a priori* SNR, R_k is the noisy spectral amplitude, $v_k = \frac{\xi_k}{1+\xi_k}\gamma_k$, and γ_k is the *a posteriori* SNR (Erkelens et al., 2007). Often this is based on a Gaussian model of noise, as it is here (Ephraim and Malah, 1985).

As mentioned, there are many utterances in which human caregivers speak concurrently with the participants. This is partially confounded by the fact that utterances by individuals with AD tend to be shorter, so more of their utterance is lost, proportionally. Examples of this type where the caregiver's voice is louder than the participant's voice are discarded, amounting to about 10% of all utterances. In the following analyses, function words (i.e., prepositions, subordinating conjunctions, and determiners) are removed from consideration, although interjections are kept. Proper names are also omitted.

We use the HTK (Young et al., 2006) toolchain, which provides an implementation of a semi-continuous hidden Markov model (HMM) that allows state-tying and represents output densities by mixtures of Gaussians. Features consisted of the first 13 Mel-frequency cepstral coefficients, their first (δ) and second ($\delta\delta$) derivatives, and the log energy component, for 42 dimensions. Our own data were z-scaled regardless of whether LSAE noise reduction was applied.

Two language models (LMs) are used, both trigram models derived from the English Gigaword corpus, which contains 1200 word tokens (Graff and Cieri, 2003). The first LM uses the first 5000 most frequent words and the second uses the first 64,000 most frequent words of that corpus. Five acoustic models (AMs) are used with 1, 2, 4, 8, and 16 Gaussians per output density respectively. These are trained with approximately 211 hours of spoken transcripts of the *Wall Street Journal* (WSJ) from over one hundred non-pathological speakers (Vertanen, 2006).

Table 2 shows, for the small- and large-vocabulary LMs, the word-level accuracies of the baseline HTK ASR system, as determined by the inverse of the Levenshtein edit distance, for two scenarios (sit-down interviews vs. during the task), with and without LSAE noise reduction, for speech from individuals with AD and for their caregivers. These values are computed over all complexities of acoustic model and are consistent with other tasks of this type (i.e., with the challenges associated with the population and recording set up), with this type of relatively unconstrained ASR (Rudzicz et al., 2012). Applying LSAE results in a significant increase in accuracy for both the small-vocabulary (right-tailed homoscedastic $t(58) = 3.9, p < 0.005, CI =$

$[6.19, \infty])$ and large-vocabulary (right-tailed homoscedastic $t(58) = 2.4, p < 0.01, CI = [2.58, \infty])$ tasks. For the participants with AD, ASR accuracy is significantly higher in interviews (paired $t(39) = 8.7, p < 0.0001, CI = [13.8, \infty])$, which is expected due in large part to the closer proximity of the microphone. Surprisingly, ASR accuracy on participants with ASR was not significantly different than on caregivers (two-tailed heteroscedastic $t(78) = -0.32, p = 0.75, CI = [-5.54, 4.0])$.

Figure 3 shows the mean ASR accuracy, with standard error ($\sigma/\sqrt{n}$), for each of the small-vocabulary and large-vocabulary ASR systems. The exponential function $b_0 + b_1 \exp(b_2 x)$ is fit to these data for each set, where b_i are coefficients that are iteratively adjustable via mean squared error. For the small-vocabulary data, $R^2 = 0.277$ and $F_8 = 3.06, p = 0.12$ versus the constant model. For the large-vocabulary data, $R^2 = 0.445$ and $F_8 = 2.81, p = 0.13$ versus the constant model. Clearly, there is an increasing trend in ASR accuracy with MMSE scores, however an n-way ANOVA on ASR accuracy scores reveals that this increase is not significant ($F_1 = 47.07, p = 0.164$). Furthermore, neither the age ($F_1 = 1.39, p = 0.247$) nor the sex ($F_1 = 0.98, p = 0.33$) of the participant had a significant effect on ASR accuracy. An additional n-way ANOVA reveals no strong interaction effects between age, sex, and MMSE.
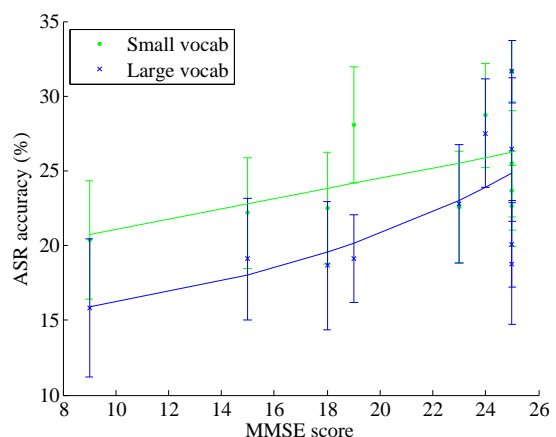

Figure 3: MMSE score versus mean ASR accuracy (with std. error bars) and fits of exponential regression for each of the small-vocabulary and large-vocabulary ASR systems.

	Scenario	Noise reduction	AD	caregiver
Small vocabulary	Interview	None	25.1 ($\sigma = 9.9$)	28.8 ($\sigma = 6.0$)
		LSAE	40.9 ($\sigma = 5.6$)	40.2 ($\sigma = 5.3$)
	In task	None	13.7 ($\sigma = 3.7$)	-
		LSAE	19.2 ($\sigma = 9.8$)	-
Large vocabulary	Interview	None	23.7 ($\sigma = 12.9$)	27.0 ($\sigma = 10.0$)
		LSAE	38.2 ($\sigma = 6.3$)	35.1 ($\sigma = 11.2$)
	In task	None	5.8 ($\sigma = 3.7$)	-
		LSAE	14.3 ($\sigma = 12.8$)	-

Table 2: ASR accuracy (means, and std. dev.) across speakers, scenario (interviews vs. during the task), and presence of noise reduction for the small and large language models.

5 Discussion

This study examined low-level aspects of speech recognition among older adults with Alzheimer's disease interacting with a robot in a simulated home environment. The best word-level accuracies of 40.9% ($\sigma = 5.6$) and 39.2% ($\sigma = 6.3$) achievable with noise reduction and in a quiet interview setting are comparable with the state-of-the-art in unrestricted large-vocabulary text entry. These results form the basis for ongoing work in ASR and interaction design for this domain. The trigram language model used in this work encapsulates the statistics of a large amount of speech from the general population – it is a speaker-independent model derived from a combination of English news agencies that is not necessarily representative of the type of language used in the home, or by our target population. The acoustic models were also derived from newswire data read by younger adults in quiet environments. We are currently training and adapting language models tuned specifically to older adults with Alzheimer's disease using data from the Carolina Conversations database (Pope and Davis, 2011) and the DementiaBank database (Boller and Becker, 1983).

Additionally, to function realistically, a lot of ambient and background noise will need to be overcome. We are currently looking into deploying a sensor network in the HomeLab that will include microphone arrays. Another method of improving rates of correct word recognition is to augment the process from redundant information from a concurrent sensory stream, i.e., in multimodal interaction (Rudzicz, 2006). Combining gesture and eye gaze with speech, for example, can be used to disambiguate speech-only signals.

Although a focus of this paper, verbal information is not the only modality in which human-robot interaction can take place. Indeed, Wilson et al. (2012) showed that experienced human caregivers employed various non-verbal and semi-verbal strategies to assist older adults with dementia about $1/3$ as often as verbal strategies (see section 2.2). These non-verbal and semi-verbal strategies included eye contact, sitting face-to-face, using hand gestures, a calm tone of voice, instrumental touch, exaggerated facial expressions, and moving slowly. Multi-modal communication can be extremely important for individuals with dementia, who may require redundant channels for disambiguating communication problems, especially if they have a language impairment or a significant hearing impairment.

It is vital that our current technological approaches to caring for the elderly in their homes progresses quickly, given the demographic shift in many nations worldwide. This paper provides a baseline assessment for the types of technical and communicative challenges that will need to be overcome in the near future to provide caregiving assistance to a growing number of older adults.

6 Acknowledgements

References

American Psychiatric Association. 2000. Delirium, dementia, and amnestic and other cognitive disorders. In *Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR)*, chapter 2. American Psychiatric Association, Arlington, VA, fourth edition.

A. Bauer, D. Wollherr, and M. Buss. 2008. Human-robot collaboration: A survey. *International Journal of Humanoid Robotics*, 5:47–66.

Momotaz Begum, Rosalie Wang, Rajibul Huq, and Alex Mihailidis. 2013. Performance of daily activities by older adults with dementia: The role of an assistive robot. In *Proceedings of the IEEE International Conference on Rehabilitation Robotics*, Washington USA, June.

Ashok J. Bharucha, Vivek Anand, Jodi Forlizzi, Mary Amanda Dew, Charles F. Reynolds III, Scott Stevens, and Howard Wactlar. 2009. Intelligent assistive technology applications to dementia care: Current capabilities, limitations, and future challenges. *American Journal of Geriatric Psychiatry*, 17(2):88–104, February.

François Boller and James Becker. 1983. Dementia-Bank database.

J.L. Burke and R.R. Murphy. 1999. Situation awareness, team communication, and task performance in robot-assisted technical search: Bujold goes to bridgeport. CMPSCI Tech. Rep. CRASAR-TR2004-23, University of South Florida.

B. Davis and M. Maclagan. 2009. Examining pauses in Alzheimer's discourse. *American journal of Alzheimer's Disease and other dementias*, 24(2):141–154.

Y. Ephraim and D. Malah. 1985. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(2):443 – 445, apr.

Jan Erkelens, Jesper Jensen, and Richard Heusdens. 2007. A data-driven approach to optimizing spectral speech enhancement methods for various error criteria. *Speech Communication*, 49:530–541.

M. F. Folstein, S. E. Folstein, T. White, and M. A. Messer. 2001. *Mini-Mental State Examination user's guide*. Odessa (FL): Psychological Assessment Resources.

A. Freedy, E. de Visser, G. Weltman, and N. Coeyman. 2007. Measurement of trust in human-robot collaboration. In *Proceedings of International Conference on Collaborative Technologies and Systems*, pages 17 –24.

Serge Gauthier, Michel Panisset, Josephine Nalbantoglu, and Judes Poirier. 1997. Alzheimer's disease: current knowledge, management and research. *Canadian Medical Association Journal*, 157:1047–1052.

R. Goldfarb and M.J.S. Pietro. 2004. Support systems: Older adults with neurogenic communication disorders. *Journal of Ambulatory Care Management*, 27(4):356–365.

M. A. Goodrich and A. C. Schultz. 2007. Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction*, 1:203–275.

David Graff and Christopher Cieri. 2003. English gigaword. Linguistic Data Consortium.

S. A. Green, M. Billinghurst, X. Chen, and J. G. Chase. 2008. Human-robot collaboration: A literature review and augmented reality approach in design. *International Journal Advanced Robotic Systems*, 5:1–18.

Curry Guinn and Anthony Habash. 2012. Technical Report FS-12-01, Association for the Advancement of Artificial Intelligence.

T Hopper. 2001. Indirect interventions to facilitate communication in Alzheimers disease. *Seminars in Speech and Language*, 22(4):305–315.

S. Klemmer, B. Hartmann, and L. Takayama. 2006. How bodies matter: five themes for interaction design. In *Proceedings of the conference on Designing Interactive systems*, pages 140–149.

Tracy Lee and Alex Mihaildis. 2005. An intelligent emergency response system: Preliminary development and testing of automated fall detection. *Journal of Telemedicine and Telecare*, 11:194–198.

Eric Lucet. 2012. Social Mobiserv Kompai Robot to Assist People. In *euRobotics workshop on Robots in Healthcare and Welfare*.

Alex Mihailidis, Jennifer N Boger, Tammy Craig, and Jesse Hoey. 2008. The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC Geriatrics*, 8(28).

Mehdi Mouad, Lounis Adouane, Pierre Schmitt, Djamel Khadraoui, Benjamin Gâteau, and Philippe Martinet. 2010. Multi-agents based system to coordinate mobile teamworking robots. In *Proceedings of the 4th Companion Robotics Institute*, Brussels.

Elizabeth D. Mynatt, Anne-Sophie Melenhorst, Arthur D. Fisk, and Wendy A. Rogers. 2004. Aware technologies for aging in place: Understanding user needs and attitudes. *IEEE Pervasive Computing*, 3:36–41.

Patrick Olivier, Andrew Monk, Guangyou Xu, and Jesse Hoey. 2009. Ambient kitchen: Designing situation services using a high fidelity prototyping environment. In *Proceedings of the ACM 2nd International Conference on Pervasive Technologies Related to Assistive Environments*, Corfu Greece.

S. V. Pakhomov, G. E. Smith, D. Chacon, Y. Feliciano, N. Graff-Radford, R. Caselli, and D. S. Knopman. 2010. Computerized analysis of speech and language to identify psycholinguistic correlates of frontotemporal lobar degeneration. *Cognitive and Behavioral Neurology*, 23:165–177.

M. E. Pollack. 2006. Autominder: A case study of assistive technology for elders with cognitive impairment. *Generations*, 30:67–69.

Charlene Pope and Boyd H. Davis. 2011. Finding a balance: The Carolinas Conversation Collection. *Corpus Linguistics and Linguistic Theory*, 7(1).

J. Reilly, J. Troche, and M. Grossman. 2011. Language processing in dementia. In A. E. Budson and N. W. Kowall, editors, *The Handbook of Alzheimer's Disease and Other Dementias*. Wiley-Blackwell.

Brian Roark, Margaret Mitchell, John-Paul Hosom, Kristy Hollingshead, and Jeffery Kaye. 2011. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2081–2090.

Elizabeth Rochon, Gloria S. Waters, and David Caplan. 2000. The Relationship Between Measures of Working Memory and Sentence Comprehension in Patients With Alzheimer's Disease. *Journal of Speech, Language, and Hearing Research*, 43:395–413.

Frank Rudzicz, Rozanne Wilson, Alex Mihailidis, Elizabeth Rochon, and Carol Leonard. 2012. Communication strategies for a computerized caregiver for individuals with alzheimer's disease. In *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies (SLPAT2012) at the 13th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2012)*, Montreal Canada, June.

Frank Rudzicz. 2006. Clavius: Bi-directional parsing for generic multimodal interaction. In *Proceedings of the joint meeting of the International Conference on Computational Linguistics and the Association for Computational Linguistics*, Sydney Australia.

Privender Saini, Boris de Ruyter, Panos Markopoulos, and Albert van Breemen. 2005. Benefits of social intelligence in home dialogue systems. In *Proceedings of INTERACT 2005*, pages 510–521.

A. Serna, H. Pigot, and V. Rialle. 2007. Modeling the progression of alzheimer's disease for cognitive assistance in smart homes. *User Modelling and User-Adapted Interaction*, 17:415–438.

Jeff A. Small, Elaine S. Andersen, and Daniel Kempler. 1997. Effects of working memory capacity on understanding rate-altered speech. *Aging, Neuropsychology, and Cognition*, 4(2):126–139.

M. Snover, B. Dorr, and R. Schwartz. 2004. A lexically-driven algorithm for disfluency detection. In *'Proceedings of HLT-NAACL 2004: Short Papers*, pages 157–160.

Adriana Tapus and Mohamed Chetouani. 2010. ROBADOM: the impact of a domestic robot on the psychological and cognitive state of the elderly with mild cognitive impairment. In *Proceedings of the International Symposium on Quality of Life Technology Intelligent Systems for Better Living*, Las Vegas USA, June.

Cheryl K. Tomoeda, Kathryn A. Bayles, Daniel R. Boone, Alfred W. Kaszniak, and Thomas J. Slauson. 1990. Speech rate and syntactic complexity effects on the auditory comprehension of alzheimer patients. *Journal of Communication Disorders*, 23(2):151 – 161.

Keith Vertanen. 2006. Baseline WSJ acoustic models for HTK and Sphinx: Training recipes and recognition experiments. Technical report, Cavendish Laboratory, University of Cambridge.

Rozanne Wilson, Elizabeth Rochon, Alex Mihailidis, and Carol Leonard. 2012. Examining success of communication strategies used by formal caregivers assisting individuals with alzheimer's disease during an activity of daily living. *Journal of Speech, Language, and Hearing Research*, 55:328–341, April.

Steve Young, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Xunying (Andrew) Liu, Gareth Moore, Julian Odell, Dave Ollason and Dan Povey, Valtcho Valtchev, and Phil Woodland. 2006. The HTK Book (version 3.4).