Communication strategies for a computerized caregiver for individuals with Alzheimer's disease

Frank Rudzicz[1,2,*] and **Rozanne Wilson**[1] and **Alex Mihailidis**[2] and **Elizabeth Rochon**[1]
[1] Department of Speech-Language Pathology,
[2] Department of Occupational Science and Occupational Therapy
University of Toronto
Toronto Canada

Carol Leonard
School of Rehabilitation Sciences
University of Ottawa
Ottawa Canada

Abstract

Currently, health care costs associated with aging at home can be prohibitive if individuals require continual or periodic supervision or assistance because of Alzheimer's disease. These costs, normally associated with human caregivers, can be mitigated to some extent given automated systems that mimic some of their functions. In this paper, we present inaugural work towards producing a generic automated system that assists individuals with Alzheimer's to complete daily tasks using verbal communication. Here, we show how to improve rates of correct speech recognition by preprocessing acoustic noise and by modifying the vocabulary according to the task. We conclude by outlining current directions of research including specialized grammars and automatic detection of confusion.

1 Introduction

In the United States, approximately $100 billion are spent annually on the direct and indirect care of individuals with Alzheimer's disease (AD), the majority of which is attributed to long-term institutional care (Ernst et al., 1997). As the population ages, the incidence of AD will double or triple, with Medicare costs alone reaching $189 billion in the US by 2015 (Bharucha et al., 2009). Given the growing need to support this population, there is an increasing interest in the design and development of technologies that support this population at home and extend ones quality of life and autonomy (Mihailidis et al., 2008).

*Contact: frank@cs.toronto.edu

Alzheimer's disease is a type of progressive neuro-degenerative dementia characterized by marked declines in mental acuity, specifically in cognitive, social, and functional capacity. A decline in memory (short- and long-term), executive capacity, visual-spacial reasoning, and linguistic ability are all typical effects of AD (Cummings, 2004). These declines make the completion of activities of daily living (e.g., finances, preparing a meal) difficult and more severe declines often necessitate caregiver assistance. Caregivers who assist individuals with AD at home are common, but are often the precursor to placement in a long-term care (LTC) facility (Gaugler et al., 2009).

We are building systems that automate, where possible, some of the support activities that currently require family or formal (i.e., employed) caregivers. Specifically, we are designing an intelligent dialog component that can engage in two-way speech communication with an individual in order to help guide that individual towards the completion of certain daily household tasks, including washing ones hands and brushing ones teeth. A typical installation setup in a bathroom, shown in figure 1, consists of video cameras that track a user's hands and the area in and around the sink, as well as microphones, speakers, and a screen that can display prompting information. Similar installations are being tested in other household rooms as part of the COACH project (Mihailidis et al., 2008), according to the task; this is an example of ambient intelligence in which technology embedded in the environment is sensitive to the activities of the user with it (Spanoudakis et al., 2010).

NAACL-HLT 2012 Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), pages 47–55,
Montréal, Canada, June 7–8, 2012.

Our goal is to encode in software the kinds of techniques used by caregivers to help their clients achieve these activities; this includes automatically identifying and recovering from breakdowns in communication and flexibly adapting to the individual over time. Before such a system can be deployed, the underlying models need to be adjusted to the desired population and tasks. Similarly, the speech output component would need to be programmed according to the vocabularies, grammars, and dialog strategies used by caregivers. This paper presents preliminary experiments towards dedicated speech recognition for such a system. Evaluation data were collected as part of a larger project examining the use of communication strategies by formal caregivers while assisting residents with moderate to severe AD during the completion of toothbrushing (Wilson et al., 2012).

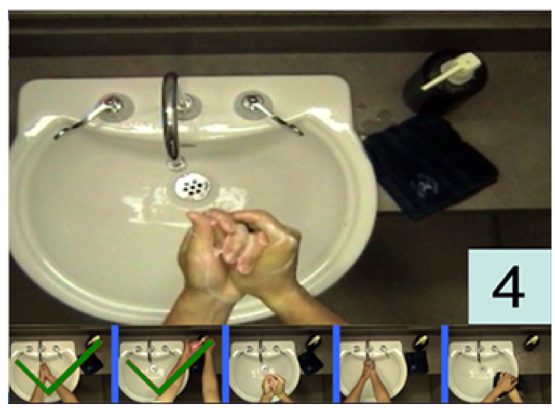
2 Background – communication strategies

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 (Small et al., 2003) but are not limited to:

1. Relatively slow rate of speech rate.
2. Verbatim repetition of misunderstood prompts.
3. Closed-ended questions (i.e., that elicit yes/no responses).
4. Simple sentences with reduced syntactic complexity.
5. Giving one question or one direction at a time.
6. Minimal use of pronouns.

These strategies, though often based on observational studies, are not necessarily based on quantitative empirical research and may not be generalizable across relevant populations. Indeed, Tomoeda et al. (1990) showed that rates of speech that are too slow may interfere with comprehension if they introduce

(a) Environmental setup



(b) On-screen prompting

Figure 1: Setup and on-screen prompting for COACH. The environment includes numerous sensors including microphones and video cameras as well as a screen upon which prompts can be displayed. In this example, the user is prompted to lather their hands after having applied soap. Images are copyright Intelligent Assistive Technology and Systems Lab).

problems of short-term retention of working memory. Small, Andersen, and Kempler (1997) showed that paraphrased repetition is just as effective as verbatim repetition (indeed, syntactic variation of common semantics may assist comprehension). Furthermore, Rochon, Waters, and Caplan (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.

Although not the empirical subject of this paper, we are studying methods of automating the resolution of communication breakdown. Much of this work is based on the Trouble Source-Repair (TSR) model in which difficulties in speaking, hearing, or understanding are identified and repairs are initiated and carried out (Schegloff, Jefferson, and Sacks, 1977). Difficulties can arise in a number of dimensions including phonological (i.e., mispronunciation), morphological/syntactic (e.g., incorrect agreement among constituents), semantic (e.g., disturbances related to lexical access, word retrieval, or word use), and discourse (i.e., misunderstanding of topic, shared knowledge, or cohesion) (Orange, Lubinsky, and Higginbotham, 1996). The majority of TSR sequences involve self-correction of a speaker's own error, e.g., by repetition, elaboration, or reduction of a troublesome utterance (Schegloff, Jefferson, and Sacks, 1977). Orange, Lubinsky, and Higginbotham (1996) showed that while 18% of non-AD dyad utterances involved TSR, whereas 23.6% of early-stage AD dyads and 33% of middle-stage AD dyads involved TSR. Of these, individuals with middle-stage AD exhibited more discourse-related difficulties including inattention, failure to track propositions and thematic information, and deficits in working memory. The most common repair initiators and repairs given communication breakdown involved frequent '*wh*-questions and hypotheses (e.g., "*Do you mean?*"). Conversational partners of individuals with middle-stage AD initiated repair less frequently than conversational partners of control subjects, possibly aware of their deteriorating ability, or to avoid possible further confusion. An alternative although very closely related paradigm for measuring communication breakdown is Trouble Indicating Behavior (TIB) in which the confused participant implicitly or explicitly requests aid. In a study of 7 seniors with moderate/severe dementia and 3 with mild/moderate dementia, Watson (1999) showed that there was a significant difference in TIB use ($\rho < 0.005$) between individuals with AD and the general population. Individuals with AD are most likely to exhibit dysfluency, lack of uptake in the dialog, metalinguistic comments (e.g., "*I can't think of the word*"), neutral requests for repetition, whereas the general population are most likely to exhibit hypothesis formation to resolve ambiguity (e.g., "*Oh, so you mean that you had a good time?*") or requests for more information.

2.1 The task of handwashing

Our current work is based on a study completed by Wilson et al. (2012) towards a systematic observational representation of communication behaviours of formal caregivers assisting individuals with moderate to severe AD during hand washing. In that study, caregivers produced 1691 utterances, 78% of which contained at least one communication strategy. On average, 23.35 ($\sigma = 14.11$) verbal strategies and 7.81 ($\sigma = 5.13$) non-verbal strategies were used per session. The five most common communication strategies employed by caregivers are ranked in table 1. The *one proposition* strategy refers to using a single direction, request, or idea in the utterance (e.g. "turn the water on"). The *closed-ended question* strategy refers to asking question with a very limited, typically binary, response (e.g., "can you turn the taps on?") as opposed to questions eliciting a more elaborate response or the inclusion of additional information. The *encouraging comments* strategy refers to any verbal praise of the resident (e.g., "you are doing a good job"). The *paraphrased repetition* strategy is the restatement of a misunderstood utterance using alternative syntactic or lexical content (e.g., "soap up your hands....please use soap on your hands"). There was no significant difference between the use of paraphrased and verbatim repetition of misunderstood utterances. Caregivers also reduced speech rate from an average baseline of 116 words per minute (s.d. 36.8) to an average of 36.5 words per minute (s.d. 19.8).

The least frequently used communication strate-

Verbal strategy	Number of occurrences Overall	Number of occurrences Successful	% use of strategy Overall	% use of strategy Successful	Uses per session Mean	Uses per session SD
One proposition	619	441	35	36	8.6	6.7
Closed-ended question	215	148	12	12	3.0	3.0
Encouraging comments	180	148	10	12	2.9	2.5
Use of resident's name	178	131	10	11	2.8	2.5
Paraphrased repetition	178	122	10	10	3.0	2.5

Table 1: Most frequent verbal communication strategies according to their number of occurrences in dyad communication. The % use of strategy is normalized across all strategies, most of which are not listed. These results are split according to the total number of uses and the number of uses in successful resolution of a communication breakdown. Mean (and standard deviation) of uses per session are given across caregivers. Adapted with permission from Wilson et al. (2012).

gies employed by experienced caregivers involved asking questions that required verification of a resident's request or response (e.g., "do you mean that you are finished?"), explanation of current actions (e.g., "I am turning on the taps for you"), and open-ended questions (e.g., "how do you wash your hands?").

The most common non-verbal strategies employed by experienced caregivers were *guided touch* (193 times, 122 of which were successful) in which the caregiver physically assists the resident in the completion of a task, *demonstrating action* (113 times. 72 of which were successful) in which an action is illustrated or mimicked by the caregiver, *handing an object to the resident* (107 times, 85 of which were successful), and *pointing to an object* (105 times, 95 of which were successful) in which the direction to an object is visually indicated by the caregiver. Some of these strategies may be employed by the proposed system; for example, videos demonstrating an action may be displayed on the screen shown in figure 1(a), which may replace to some extent the mimicry by the caregiver. A possible replication of the fourth most common non-verbal strategy may be to highlight the required object with a flashing light, a spotlight, or by displaying it on screen; these solutions require tangential technologies that are beyond the scope of this current study, however.

3 Data

Our experiments are based on data collected by Wilson et al. (submitted) with individuals diagnosed with moderate-to-severe AD who were recruited from long-term care facilities (i.e., The Harold and Grace Baker Centre and the Lakeside Long-Term Care Centre) in Toronto. Participants had no previous history of stroke, depression, psychosis, alcoholism, drug abuse, or physical aggression towards caregivers. Updated measures of disease severity were taken according to the Mini-Mental State Examination (Folstein, Folstein, and McHugh, 1975). The average cognitive impairment among 7 individuals classified as having severe AD (scores below $10/30$) was 3.43 ($\sigma = 3.36$) and among 6 individuals classified as having moderate AD (scores between $10/30$ and $19/30$) was 15.8 ($\sigma = 4.07$). The average age of residents was 81.4 years with an average of 13.8 years of education and 3.1 years of residency at their respective LTC facility. Fifteen formal caregivers participated in this study and were paired with the residents (i.e., as dyads) during the completion of activities of daily living. All but one caregiver were female and were comfortable with English. The average number of years of experience working with AD patients was 12.87 ($\sigma = 9.61$).

The toothbrushing task follows the protocol of the handwashing task. In total, the data consists of 336 utterances by the residents and 2623 utterances by their caregivers; this is manifested by residents uttering 1012 words and caregivers uttering 12166 words in total, using 747 unique terms. The toothbrushing task consists of 9 subtasks, namely: 1) get brush and paste, 2) put paste on brush, 3) turn on water, 4) wet tooth brush, 5) brush teeth, 6) rinse mouth, 7) rinse brush, 8) turn off water, 9) dry mouth.

These data were recorded as part of a large project to study communication strategies of caregivers rather than to study the acoustics of their transactions with residents. As a result, the recordings were not of the highest acoustic quality; for example, although the sampling rate and bit rate were high (48 kHz and 384 kbps respectively), the video camera used was placed relatively far from the speakers, who generally faced away from the microphone towards the sink and running water. The distribution of strategies employed by caregivers for this task is the subject of ongoing work.

4 Experiments in speech recognition

Our first component of an automated caregiver is the speech recognition subsystem. We test two alternative systems, namely Carnegie Mellon's Sphinx framework and Microsoft's Speech Platform. Carnegie Mellon's Sphinx framework (pocketsphinx, specifically) is an open-source speech recognition system that uses traditional N-gram language modeling, sub-phonetic acoustic hidden Markov models (HMMs), Viterbi decoding and lexical-tree structures (Lamere et al., 2003). Sphinx includes tools to perform traditional Baum-Welch estimation of acoustic models, but there were not enough data for this purpose. The second ASR system, Microsoft's Speech Platform (version 11) is less open but exposes the ability to vary the lexicon, grammar, and semantics. Traditionally, Microsoft has used continuous-density HMMs with 6000 tied HMM states (senones), 20 Gaussians per state, and Mel-cepstrum features (with delta and delta-delta).

Given the toothbrushing data described in section 3, two sets of experiments were devised to configure these systems to the task. Specifically, we perform preprocessing of the acoustics to remove environmental noise associated with toothbrushing and adapt the lexica of the two systems, as described in the following subsections.

4.1 Noise reduction

An emergent feature of the toothbrushing data is very high levels of acoustic noise caused by the running of water. In fact, the estimated signal-to-noise ratio across utterances range from -2.103 dB to 7.63 dB, which is extremely low; for comparison clean speech typically has an SNR of approximately 40 dB. Since the resident is likely to be situated close to this source of the acoustic noise, it becomes important to isolate their speech in the incoming signal.

Speech enhancement involves the removal of acoustic noise $d(t)$ in a signal $y(t)$, including ambient noise (e.g., running water, wind) and signal degradation giving the clean 'source' signal $x(t)$. This involves an assumption that noise is strictly additive, as in the formula:

$$y(t) = x(t) + d(t). \tag{1}$$

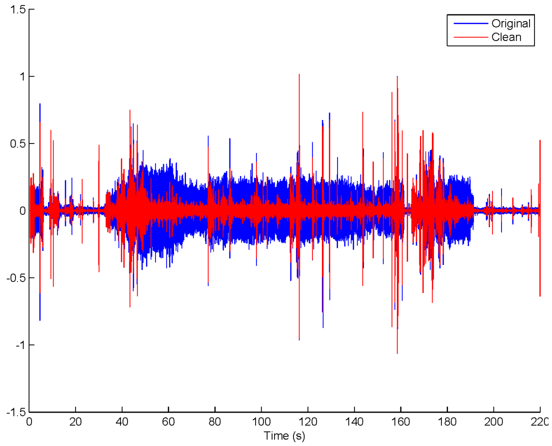
Here, Y_k, X_k, and D_k are the k^{th} spectra of the noisy observation $y(t)$, source signal $x(t)$, and uncorrelated noise signal $d(t)$, respectively. Generally, the spectral magnitude of a signal is more important than its phase when assessing signal quality and performing speech enhancement. Spectral subtraction (SS), as the name suggests, subtracts an estimate of the noisy spectrum from the measured signal (Boll, 1979; Martin, 2001), where the estimate of the noisy signal is estimated from samples of the noise source exclusively. That is, one has to learn estimates based on pre-selected recordings of noise. We apply SS speech enhancement given sample recordings of water running. The second method of enhancement we consider is the log-spectral amplitude estimator (LSAE) which minimizes the mean squared error (MMSE) of the log spectra given a model for the source speech $X_k = A_k \exp(j\omega_k)$, where A_k is the spectral amplitude. The LSAE method is a modification to the short-time spectral amplitude estimator that attempts to find some estimate $\hat{A}_k$ that minimizes the distortion

$$E\left[\left(logA_k - \log \hat{A}_k\right)^2\right], \tag{2}$$

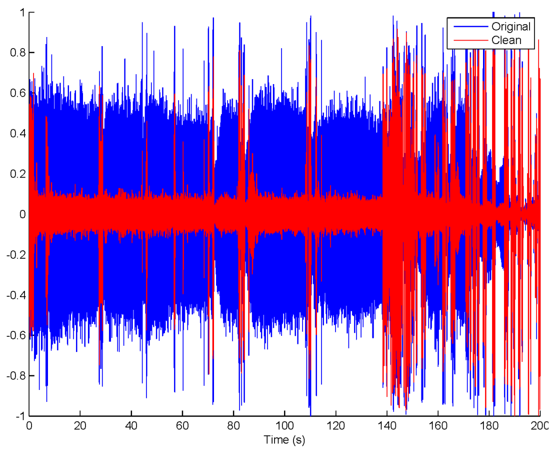
such that the log-spectral amplitude estimate is

$$\begin{aligned}\hat{A}_k &= \exp\left(E\left[\ln A_k \mid Y_k\right]\right)\\ &= \frac{\xi_k}{1+\xi_k}\exp\left(\frac{1}{2}\int_{v_k}^{\infty}\frac{e^{-t}}{t}dt\right)R_k,\end{aligned} \tag{3}$$

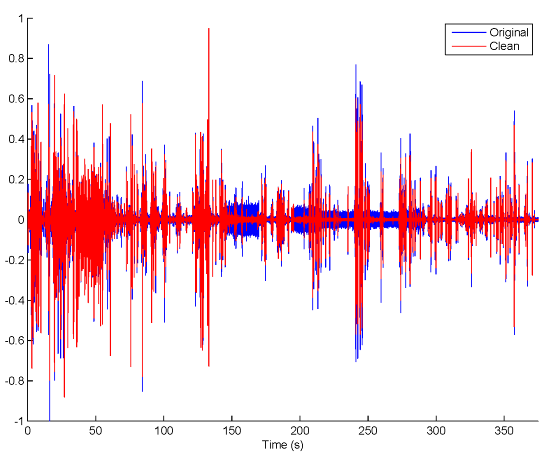
where ξ_k is the *a priori* SNR, R_k is the noisy spectral amplitude, $v_k = \frac{\xi_k}{1+\xi_k}\gamma_k$, and γ_k is the *a posteriori* SNR (Erkelens, Jensen, and Heusdens, 2007). Often this is based on a Gaussian model of noise, as it is here (Ephraim and Malah, 1985). We enhance our recordings by both the SS and LSAE methods. Archetypal instances of typical, low, and (relatively) high SNR waveform recordings and their enhanced versions are shown in 4.1.



(a) Dyad1.1



(b) Dyad4.2



(c) Dyad11.1

Figure 2: Representative samples of toothbrushing data audio. Figures show normalized amplitude over time for signals cleaned by the LSAE method overlaid over the larger-amplitude original signals.

We compare the effects of this enhanced audio across two ASR systems. For the Sphinx system, we use a continuous tristate HMM for each of the 40 phones from the CMU dictionary trained with audio from the complete Wall Street Journal corpus and the independent variable we changed was the number of Gaussians per state (n. Γ). These parameters are not exposed by the Microsoft speech system, so we instead vary the minimum threshold of confidence $\mathcal{C} \in [0..1]$ required to accept a word; in theory lower values of $\mathcal{C}$ would result in more insertion errors and higher values would result in more deletion errors. For each system, we used a common dictionary of $123,611$ unique words derived from the Carnegie Mellon phonemic dictionary.

Table 2 shows the word error rate for each of the two systems. Both the SS and LSAE methods of speech enhancement result in significantly better word error rates than with the original recordings at the 99.9% level of confidence according to the one-tailed paired t-test across both systems. The LSAE method has significantly better word error rates than the SS method at the 99% level of confidence with this test. Although these high WERs are impractical for a typical system, they are comparable to other results for speech recognition in very low-SNR environments (Kim and Rose, 2003). Deng et al. (2000), for example, describe an ASR system trained with clean speech that has a WER of 87.11% given additive white noise for a resulting 5 dB SNR signal for a comparable vocabulary of 5000 words. An interesting observation is that even at the low confidence threshold of $\mathcal{C} = 0.2$, the number of insertion errors did not increase dramatically relative to for the higher values in the Microsoft system; only 4.0% of all word errors were insertion errors at $\mathcal{C} = 0.2$, and 2.7% of all word errors at $\mathcal{C} = 0.8$.

Given Levenshtein alignments between annotated target (reference) and hypothesis word sequences, we separate word errors across residents and across caregivers. Specifically, table 3 shows the proportion of deletion and substitution word errors (relative to totals for each system separately) across residents and caregivers. This analysis aims to uncover differences in rates of recognition between those with AD and the more general population. For example, 12.6% of deletion errors made by Sphinx were words spoken by residents. It is not possible to at-

	Parameters	Word error rate % Original	SS	LSAE
Sphinx	n. $\Gamma = 4$	98.13	75.31	70.61
	n. $\Gamma = 8$	98.13	74.95	69.66
	n. $\Gamma = 16$	97.82	75.09	69.78
	n. $\Gamma = 32$	97.13	74.88	67.22
Microsoft	$\mathcal{C} = 0.8$	97.67	73.59	67.11
	$\mathcal{C} = 0.6$	97.44	72.57	67.08
	$\mathcal{C} = 0.4$	96.85	71.78	66.54
	$\mathcal{C} = 0.2$	94.30	71.36	64.32

Table 2: Word error rates for the Sphinx and Microsoft ASR systems according to their respective adjusted parameters, i.e., number of Gaussians per HMM state (n. Γ) and minimum confidence threshold ($\mathcal{C}$). Results are given on original recordings and waveforms enhanced by spectral subraction (SS) and MMSE with log-spectral amplitude estimates (LSAE).

tribute word insertion errors to either the resident or caregiver, in general. If we assume that errors should be distributed across residents and caregivers in the same proportion as their respective total number of words uttered, then we can compute the Pearson χ^2 statistic of significance. Given that 7.68% of all words were uttered by residents, the observed number of substitutions was significantly different than the expected value at the 99% level of confidence for both the Sphinx and Microsoft systems, but the number of deletions was not significantly different even at the 95% level of confidence. In either case, however, substantially more errors are made proportionally by residents than we might expect; this may in part be caused by their relatively soft speech.

	Proportion of errors: Sphinx Res.	Sphinx Careg.	Microsoft Res.	Microsoft Careg.
deletion	13.9	86.1	12.6	87.4
substitution	23.2	76.8	18.4	81.6

Table 3: Proportion of deletion and substitution errors made by both (Res)idents and (Careg)ivers. Proportions are relative to totals within each system.

4.2 Task-specific vocabulary

We limit the common vocabulary used in each speech recognizer in order to be more specific to the task. Specifically, we begin with the 747 words uttered in the data as our most restricted vocabulary. Then, we expand this vocabulary according to two methods. The first method adds words that are semantically similar to those already present. This is performed by taking the most common sense for each noun, verb, adjective, and adverb, then adding each entry in the respective synonym sets according to WordNet 3.0 (Miller, 1995). This results in a vocabulary of 2890 words. At this point, we iteratively add increments of words at intervals of $10,000$ (up to $120,000$) by selecting random words in the vocabulary and adding synonym sets for all senses as well as antonyms, hypernyms, hyponyms, meronyms, and holonyms. The result is a vocabulary whose semantic domain becomes increasingly generic. The second approach to adjusting the vocabulary size is to add phonemic foils to more restricted vocabularies. Specifically, as before, we begin with the restricted 747 words observed in the data but then add increments of new words that are phonemically similar to existing words. This is done exhaustively by selecting a random word and searching for minimal phonemic misalignments (i.e., edit distance) among out-of-vocabulary words in the Carnegie Mellon phonemic dictionary. This approach of adding decoy words is an attempt to model increasing generalization of the systems. Every vocabulary is translated into the format expected by each recognizer so that each test involves a common set of words.

Word error rates are measured for each vocabulary size across each ASR system and the manner in which those vocabularies were constructed (semantic or phonemic expansion). The results are shown in figure 4.2 and are based on acoustics enhanced by the LSAE method. Somewhat surprisingly, the method used to alter the vocabulary did appear to have a very large effect. Indeed, the WER across the semantic and phonemic methods were correlated at $\rho >= 0.99$ across both ASR systems; there was no significant difference between traces (within system) even at the 60% level of confidence using the two-tailed heteroscedastic t-test.

5 Ongoing work

This work represents the first phase of development towards a complete communicative artificial caregiver for the home. Here, we are focusing on the

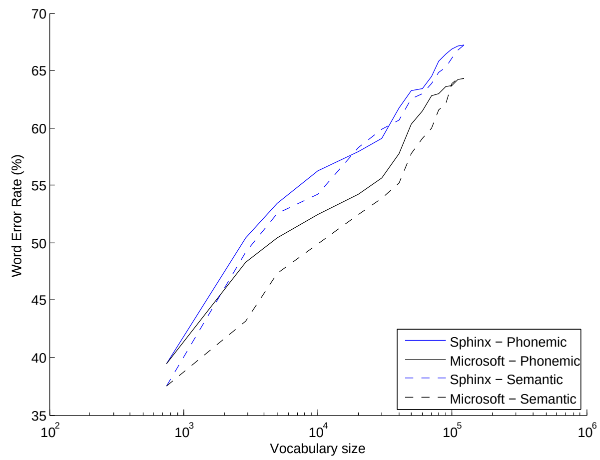


Figure 3: Word error rate versus size of vocabulary (log scale) for each of the Sphinx and Microsoft ASR systems according to whether the vocabularies were expanded by semantic or phonemic similarity.

speech recognition component and have shown reductions in error of up to 72% (Sphinx ASR with $n.\Gamma = 4$) and 63.1% (Sphinx ASR), relative to baseline rates of error. While significant, baseline errors were so severe that other techniques will need to be explored. We are now collecting additional data by fixing the Microsoft Kinect sensor in the environment, facing the resident; this is the default configuration and may overcome some of the obstacles present in our data. Specifically, the beamforming capabilities in the Kinect (generalizable to other multi-microphone arrays) can isolate speech events from ambient environmental noise (Balan and Rosca, 2002). We are also collecting speech data for a separate study in which individuals with AD are placed before directional microphones and complete tasks related to the perception of emotion.

As tasks can be broken down into non-linear (partially ordered) sets of subtasks (e.g., replacing the toothbrush is a subtask of toothbrushing), we are specifying grammars 'by hand' specific to those subtasks. Only some subset of all subtasks are possible at any given time; e.g., one can only place toothpaste on the brush once both items have been retrieved. The possibility of these subtasks depend on the state of the world which can only be estimated through imperfect techniques – typically computer vision. Given the uncertainty of the state of the world, we are integrating subtask-specific grammars into a partially-observable Markov decision process (POMDP). These grammars include the semantic state variables of the world and break each task down into a graph-structure of interdependent actions. Each 'action' is associated with its own grammar subset of words and phrases that are likely to be uttered during its performance, as well as a set of prompts to be spoken by the system to aid the user. Along these lines, we we will attempt to generalize the approach taken in section 4.2 to generate specific sub-vocabularies automatically for each subtask. The relative weighting of words will be modeled based on ongoing data collection.

Acknowledgments

This research was partially funded by Mitacs and an operating grant from the Canadian Institutes of Health Research and the American Alzheimer Association (ETAC program). The authors acknowledge and thank the administrative staff, caregivers, and residents at the Harold and Grace Baker Centre and the Lakeside Long-Term Care Centre.

References

Balan, Radu and Justinian Rosca. 2002. Microphone Array Speech Enhancement by Bayesian Estimation of Spectral Amplitude and Phase. In *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop*.

Bharucha, Ashok J., 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.

Boll, S.F. 1979. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2):113–120, April.

Cummings, Jeffrey L. 2004. Alzheimer's disease. *New England Journal of Medicine*, 351(1):56–67.

Deng, Li, Alex Acero, M. Plumpe, and Xuedong Huang. 2000. Large-vocabulary speech recognition under adverse acoustic environments. In *Proceedings of the International Conference on Spoken Language Processing*, October.

Ephraim, Y. 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.

Erkelens, Jan, 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.

Ernst, Richard L., Joel W. Hay, Catharine Fenn, Jared Tinklenberg, and Jerome A. Yesavage. 1997. Cognitive function and the costs of alzheimer disease – an exploratory study. *Archives of Neurology*, 54:687–693.

Folstein, Marshal F., Susan E. Folstein, and Paul R. McHugh. 1975. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, November.

Gaugler, J. E., F. Yu, K. Krichbaum, and J.F. Wyman. 2009. Predictors of nursing home admission for persons with dementia. *Medical Care*, 47(2):191–198.

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

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

Kim, Hong Kook and Richard C. Rose. 2003. Cepstrum-Domain Acoustic Feature Compensation Based on Decomposition of Speech and Noise for ASR in Noisy Environments. *IEEE Transactions on Speech and Audio Processing*, 11(5), September.

Lamere, Paul, Philip Kwok, Evandro Gouvea, Bhiksha Raj, Rita Singh, William Walker, M. Warmuth, and Peter Wolf. 2003. The CMU Sphinx-4 speech recognition system. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, Hong Kong, April.

Martin, Rainer. 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions of Speech and Audio Processing*, 9(5):504–512, July.

Mihailidis, Alex, 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).

Miller, George A. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41.

Orange, J.B., Rosemary B. Lubinsky, and D. Jeffery Higginbotham. 1996. Conversational repair by individuals with dementia of the alzheimer's type. *Journal of Speech and Hearing Research*, 39:881–895, August.

Rochon, Elizabeth, 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.

Saini, Privender, 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.

Schegloff, Emanuel A., Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *1977*, 53(2):361–382.

Small, Jeff A., 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.

Small, Jeff A., Gloria Gutman, Saskia Makela, and Beth Hillhouse. 2003. Effectiveness of communication strategies used by caregivers of persons with alzheimer's disease during activities of daily living. *Journal of Speech, Language, and Hearing Research*, 46(2):353–367.

Spanoudakis, Nikolaos, Boris Grabner, Christina Kotsiopoulou, Olga Lymperopoulou, Verena Moser-Siegmeth, Stylianos Pantelopoulos, Paraskevi Sakka, and Pavlos Moraitis. 2010. A novel architecture and process for ambient assisted living - the hera approach. In *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB)*, pages 1–4.

Tomoeda, Cheryl K., 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.

Watson, Caroline M. 1999. An analysis of trouble and repair in the natural conversations of people with dementia of the Alzheimer's type. *Aphasiology*, 13(3):195 – 218.

Wilson, Rozanne, Elizabeth Rochon, Alex Mihailidis, and Carol Léonard. 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.

Wilson, Rozanne, Elizabeth Rochon, Alex Mihailidis, and Carol Léonard. submitted. Quantitative analysis of formal caregivers' use of communication strategies while assisting individuals with moderate and severe alzheimer's disease during oral care. *Journal of Speech, Language, and Hearing Research*.