Speech-based interaction in an AAL context

Michel Vacher, François Portet, Solange Rossato, Frédéric Aman, Caroline Golanski and Remus Dugheanu *

LIG, GETALP Team, UMR CNRS/UJF/Grenoble-INP 5217, Grenoble, France
** Corresponding author (Michel.Vacher@imag.fr)*

One of the biggest challenges in Ambient Assisted Living is to develop smart homes that anticipate the health needs of their inhabitants while maintaining their safety and comfort. To facilitate interactions with the smart home, systems that respond naturally to voice commands would be the most adequate for disabled and frail people. In this paper we present two studies aiming at investigating the feasibility of such interactive systems. In the first study, the acceptability of a voice interface as part of the smart home was investigated. The second study is related to the adaptation of speech recognition technologies to the senior population; population which is known to challenge standard ASR systems. To this aim, we recorded two specific speech corpora (Voice-Age and ERES38) which were analyzed in a semi automatic manner to reveal the aged-voice characteristics. Some phonemes are more affected by age than others and this study shows that, by adapting acoustic models to seniors, performances increase. Voice interfaces appear to have a great potential to ease daily living for elderly and frail persons and would be better accepted than more intrusive solutions. An interesting finding that came up is their overall acceptance provided the system does not drive them to a lazy lifestyle by taking control of everything. Smart homes must support daily living by giving seniors more ability to control rather than acting in place of people.

Keywords*: voice interface, speech recognition, ageing voice, AAL, smart Home*

INTRODUCTION

The demographic change and ageing in the developed countries imply challenges in the way the population will be cared for in the near future. A popular solution is to develop ICT to make it possible for older person to stay at home. Home support of older adults is related to several constraints: first, the increasing number of older persons who often wish to live independently as long as possible in their own homes; secondly, the cost challenge to society to support people loosing independence and; thirdly, the shortage of places in specialized institutions. Given these demographic trends and using new technologies, Ambient Assisted Living (AAL) research programs intend to develop services and innovative products that improve the quality of life of older people, maintain their independence and quality of life in a normal living environment whereas seniors are more affected by physical or cognitive diseases.

Solutions must be developed to compensate the possible physical or mental decline to keep older persons with a good degree of autonomy. The aim is also to provide assistance if necessary through surveillance to detect distress situations. However, in the most general case, these persons are often confused by complex interfaces and technological solutions must be adapted to the needs and the specific capacities of this population. Therefore, as voice is the most natural way of communication, interfaces using a system of Automatic Speech Recognition (ASR) may be more accessible.

The aim of this paper is to expose the possible uses of audio and speech analysis in the AAL context before presenting two studies involving older persons.

The first study aims at evaluating how ambient assistive speech technology is received by the targeted population. We report a user evaluation assessing the acceptance and the fear of this new technology.

The second study is related to the adaptation of speech recognition technologies to the older population which has not been extensively studied in the literature, even if it is known that system development for other categories of population is not adapted to the senior. This article will terminate by a brief discussion about the results and by a presentation of the perspectives.

AUDIO USE IN AN AMBIENT ASSISTED LIVING CONTEXT

Smart Home and AAL

One of the biggest challenges in Ambient Assisted Living (AAL) is to develop *smart homes* that anticipate and respond to the needs of their inhabitants, this is especially important when they have disabilities. Smart homes are habitations equipped with a home automation system including a set of sensors, automated devices and centralized software which control the increasing amount of household appliances. Given the diverse profile of the elderly population, it is thus essential to facilitate interaction with the smart home through systems that respond naturally to voice commands rather than using tactile

interfaces which require physical and visual interaction. Therefore, although speech interaction is rarely considered, it seems more adapted to people who have difficulties in moving or seeing[1].
Chan et al.[2] reviewed the projects with a medical perspective and identified the necessary conditions to satisfy, namely: -*user needs, acceptability and satisfaction*; -*viability and efficacy* of sensors and software; -*standard compliance* for information and communication systems; -*legacy and ethical constraints*; -*cost reduction and socio-economical impact*. Moreover, several studies were conducted to identify the needs of older people towards a system that can help them in their everyday life[3,4].
The proposed systems consider providing assistance in three main areas: - **health** (tracking the status of the person and the evolution of his loss of autonomy by using physiological sensors, motion sensors, video cameras, home intrusion, etc.); - **security** (preventing and detecting situations of distress or risks through fall detectors, smoke detectors, etc.); - and **assistance** for home automation systems (compensation of disabilities through better access to domestic appliances). It should be noted that a fourth priority must be added, it is the **communication with relatives and with external persons** which is essential for the person isolated at home.

Audio sensing technology in Smart Home
Audio analysis can be divided into speech recognition and sound identification. In this context, speech recognition is useful for voice command and dialogue while sound identification gives information about the activity of the person (e.g., closing the door or using the vacuum). Many challenges have to be overcome before these technologies could be usable and deployable in assisted living applications[5]. Sounds are produced by very varied audio sources and then sound identification is less advanced than speech recognition which benefits from continuous progress since the use of probabilistic models (Hidden Markov Model-HMM) for phoneme modeling[6]. Therefore, this section focuses on speech recognition application in smart homes.

Extraction of speech in a noisy environment
In real conditions, audio processing is affected by: background noise (TV, devices, traffic…), the room acoustic (reverberation on windows) and the position of the speaker with respect to the microphones set in the room. The most significant problem is related to speech signal mixed with unwanted noises such as music or vacuum. Sophisticated signal processing techniques should be considered to solve this problem (i.e. echo cancellation, blind source separation…).

Voice interface for compensation and comfort
Voice interface is a natural way of providing the ability of driving verbally the home automation system. It's a natural way to enable a physically disabled person (e.g., person in a wheel-chair, blinds…) to keep control of their environment.

Detection of distress situations
Identifying specific sounds (glass breaking or falls) would be of great interest for such situation detection. On the other hand, voice interfaces make it possible to call for help when the person is in a distress situation but remains conscious. Moreover, gerontologists and physicians pointed out the importance of emotion in the voice[7]. Automatic emotion level recognition would be highly helpful to detect an important problem to solve and then assistance is requested.

Evaluation of the ability to communicate
One of the most tragic symptoms of Alzheimer's disease is the progressive loss of vocabulary and communication skills. The changes can be very slow and difficult to detect by caregivers, and an automatic monitoring could allow the detection of important stages of this evolution[19].

Speech recognition adapted to the speaker
Several experiments established that automatic speech recognition performances are degraded for atypical population like children or aged persons[8]. As AAL concern a lot older persons, this point will be the matter of a following section.

Privacy and acceptability
Due to the increasing number of sophisticated electronic devices, the question of privacy becomes crucial[9,10] and speech recognition must respect privacy. Therefore, the Automatic Speech Recognizer (ASR) must be adapted to the application and should not be able to recognize private conversation sentences.
Regarding the acceptability point of view, a system will be more accepted if it is regularly used in the daily life such as with a home automation system, than if it is used in rare circumstances (e.g., fall). Acceptability is crucial because it conditions the use of the system by the person. Therefore the next section is devoted to this aspect.

ACCEPTABILITY OF VOCAL ORDERS
Older persons are the most likely persons to develop cognitive and physical decline but that does not imply they are all not self-governing. Thus, the design of new daily living support technologies must take into account studies which showed that the reduction of sense of control in the elderly population may have a significant adverse effect on their health[22].

Moreover, given the implication of the relatives in supporting their seniors, caregivers must also be included in the design. To find out what the needs of this target population are, we conducted a user evaluation assessing the acceptance and the fear of the audio technology[1]. In the experiment we assessed the acceptability of such technology. This is the key factor for integrating new technologies in homes, particularly when the users are older persons or low ICT educated persons. In the following we present the experiment we set up to test users' acceptance and a summary of the results.

Experimental set-up

The experience we undertook aimed at testing three important aspects of speech interaction in a smart home: voice orders, communication with the outside world, home automation system interrupting a person's activity. It consisted in Wizard of Oz phases and interview phases in the smart home. The WOZ interaction consisted mainly in the control of the environment. For instance, if the participant said “close the blind”, the blind were closed remotely.

The participants consisted of 18 persons from the Grenoble area. Among them, 8 were in the senior group, 7 in the relatives group (composed of mature children, grandchildren or friends). The mean age of the elderly group was 79.0 (SD=6.0), and 5 out of 8 were women. These persons were single and perfectly autonomous. In order to acquire another view about the interest and acceptability of the project system, 3 professional caregivers were also recruited to participate in the experiment (2 nurses and one professional elderly assistant).

Fig.1. Each test was composed of an interviewer, a wizard and a couple of one elderly person with a relative (except for one senior who was alone) inside the DOMUS smart home of the LIG Laboratory[1]

The co-discovery approach (see Figure 1) was chosen to reassure the senior about the experimental context (new environment, experimenter, etc.) thanks to the presence of their relative. Moreover, it eased the projection of both participants into the new system because they could exchange points of view. Of course, the relationship between the two people can also influence the experiment that is why some short periods were planned during which the participants were interrogated separately.

To assess the acceptability of the system which has no standard definition, most of the experiments were conducted to find out whether the potential users would appreciate the new functionalities brought by the system (e.g., “Do you appreciate making the system operate using your voice? Why?”). Moreover, in order to guide the development of the system, aspects of usefulness, usability, interactiveness, proactiveness, intrusiveness, social interaction, and security were investigated.

The first phase of the experiment was about the voice command aspect of the project. Both the senior and her relative were present in the room. The senior was asked to control blinds, lights and the coffee machine using her voice without any recommendation about how to do it. This consisted in talking “to the home”. The second phase consisted in using technology for communication with the outside such as video conferencing. The senior was left alone in the smart home watching a TV program, when the program interrupted itself to let the face of the relative appear on the screen so that they can start a conversation. The third phase focused on system interruption. The couple and the interviewer were discussing in the smart home when the system interrupted them via a pre-recorded voice played through the speakers, calling for a door to be closed or the cooker to be turned off. After this, questions related to whether being interrupted by the system was acceptable or not. Also, the problem of security in general and how such system could enhance security was discussed with the couple.

Results of the study

From the results of the study, it appears that seniors and their relatives preferred mostly the voice command, the system interventions about safety issues and the video-conferencing. It is interesting to note that, in our study, the “key-word” form for commands is highly accepted (rather than the sentence based command). This highly simplifies the integration of such technology in smart home given that small vocabulary systems are generally performing better in real world applications than large vocabulary ones.

As in other related studies[3], all participants found a strong interest in the voice interaction system. It is strongly preferred over a tactile system (or touch-screen) which would necessitate being physically available at the place the command is to be found or would imply to constantly know where the remote controller is. This is in line with other studies concerning personal emergency response systems

which showed that push-button based control is not adapted to this population[23].

Although the system was well received, it turned out that some functionality provoked strong objections among the participants. The main fear of the elderly and relatives is the system failure. Another main concern about the system is the fact that too much assistance would increase the dependence of the person by pushing her toward inactivity. Regarding the caregivers, they expressed the concern that such system would tend to gradually replace some of their visits and end up in making the seniors even more isolated. Most of these fears can be addressed by a good design of the system. However, fear about a decrease in autonomy due to a system that can do everything is a subtle one. A system designed for active people in order to improve comfort, security and save time may not be adapted to healthy but aged persons[24].

While the proposed system can bring more comfort and autonomy to daily life by providing an easy interaction with the home automation elements, the majority of the participants insisted on the security aspects. For instance, the voice interface would be of great use in case of falls. The elderly and their relatives have particularly appreciated that the system spares the elderly actions that can be dangerous and warns them of dangerous situations. This trend is confirmed in almost all user evaluations involving elderly[25,3,4]. Thus, smart homes for the elderly would be much more accepted if they contain features that can reassure them regarding security more than any other features whatever their initial condition and origin in developed countries are.

Overall, the participants mainly stressed the interest of voice command and how this could improve security, autonomy and, to a smaller extend, could fight loneliness. However, they were very careful about privacy and clearly showed that they were very cautious of not accepting systems that would push them into a dependent situation. They want to keep control. Although only a small sample of seniors and relatives in healthy condition was recruited, this qualitative study confirmed the interest of voice-based technology for smart home and uncovered some pitfalls to avoid in its design. In the next section we describe a preliminary study to adapt standard speech recognition to the ageing voice.

AGEING VOICE

Speech of the older persons is characterized by tremor of the voice, an imprecise production of consonants, and a slower articulation[11]. From an anatomical point of view, studies have shown age-related degeneration with atrophy of the vocal cords, calcification of the laryngeal cartilages, and changes in the muscles of the larynx[12,13]. Given that most acoustic models of ASR systems are trained from non-aged voice samples, they are not adequate to deal with more aged population and then classical ASR systems exhibit poor performances[14,15,16].

To improve the acoustic-phonetic decoding module in ASR systems and to adapt it to the voice of seniors, we studied the phonemes that were poorly recognized in an aged voice. To do so, we collected corpora, used ASR alignment and analyzed the most difficult phonemes. Then, acoustic model adaptation was done in order to evaluate the ASR. The main steps of the study are summed up in Figure 1.

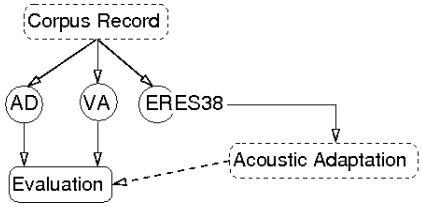

Fig.1. Main steps of the Ageing Voice study

Speech corpora

Given the absence of oral corpus adapted to the assistance at home, we recorded two new corpora specific to home automation and distress detection: the *Anodin-Détresse* (**AD** or Colloquial-Distress) corpus[17] and *Voice-Age* (**VA**) corpus[18].

AD is a corpus made of short sentences read by 21 speakers. Each of the 37 older speakers of VA read the 126 sentences of AD and long sentences extracted from newspapers or magazines.

Given the results of our first studies[18], we recorded a new corpus named **ERES38** in order to improve the acoustic models for older persons by favoring the problematic phonemes such as plosives and fricatives. The ERES38 corpus is a collection of interviews useful to collect informal and spontaneous speech recorded in specific establishments, such as nursing homes. A reading was also performed during this interview. Plosives and fricatives were introduced in the text in order to be in context /a/, /i/ and /u/. The interviews were conducted with people more or less autonomous, cognitively intact, sometimes with serious mobility problems, but without other severe disabilities. The recordings began to be transcribed, and all readings were transcribed and checked. This corpus is used for acoustic model training.

Corpus	Gender M/F	Age min-max	Duration	Sentence number
AD	10/11	20-65	38min	2 646
VA	11/26	62-94	5h19min	9 052
ERES38	8/16	68-98	17h44min	7 600

Table 1. Main characteristics of the 3 speech corpora

Influence of ageing voice on Automatic Speech Recognition

To compare the influence of the *Aged Speech* (VA) and the *Non-Aged Speech* (AD) groups we use the Sphinx3 engine as ASR[20]. Acoustic models were trained using the BREF120 corpus[21]. The language model was trained with the transcript of AD in order to match the context of home automation voice commands. The result is a very small trigram language model with a vocabulary of about 170 words focused on error analysis of the acoustic-phonetic decoding step. A more detailed description of this system is available[18].

The decoding with sphinx3 generates an orthographic transcription of the audio signal of speech. We obtained a Word Error Rate (WER) of 7.33% for the *Non-Aged Speech* group (21 speakers) and a WER of 27.56% for the decoding of the *Aged Speech* group (36 speakers). Thus we observed a significant performance degradation of ASR for elderly speech, with an absolute difference of 20.23%.

A more precise analyze is given by the forced alignment scores by phoneme. Forced alignment scores are likelihood scores belonging to the phoneme normally delivered for the considered signal portion. This score can be interpreted as a proximity to the "standard" pronunciation modeled by the generic acoustic model. The difference of acoustic score for Aged versus Non-Aged is shown in Table 2.

Phonemic group	Score difference
Nasal vowels	-117.00%
Unvoiced fricative consonants	-110.56%
Unvoiced plosive consonants	-105.72%
Voiced fricative consonants	-87.86%
Voiced plosive consonants	-83.29%
Medium vowels	-63.74%
Open vowels	-53.21%
Closed vowels	-45.52%
Nasal and liquid consonants	-42.65%

Table 2. Difference of acoustic score in forced alignment for *Aged Speech* vs. *Non-Aged Speech*

Consonants are generally most affected and the absence of voicing is the main factor of degradation followed by the modality of implementation (fricative vs. plosive). Therefore, it is possible that unvoiced consonants for older persons are closer to voiced consonants. These findings are in line with other results[16] apart for the nasal and liquid consonants and the nasal vowels where their results are poorer.

Acoustic adaptation

The adjustment method of Maximum Likelihood Linear Regression (MLLR) was used to adapt the generic acoustic model trained with BREF120 to the voice of seniors. The adaptation was made globally with all sentences of the ERES38 corpus. The new acoustic model is the MLLR adapted model. The WER was then obtained for the 36 speakers (VA) using the generic model (WER_1) and using the MLLR adapted model (WER_2).

Results of the study

The speakers are grouped together using k-means clustering method based on observations given by WER_1 and WER_2. The main characteristics of the 3 groups, number, gender and age, are reported in Table 2.

As shown in Table 3, using the MLLR Adapted Acoustic Model reduces the WER significantly up to 11.95%. Compared to the 27.56% WER without adaptation, the absolute difference is -15.61% (relative difference of -56.65%). From an applicative point of view, this shows that we can use a database of elderly speech to make MLLR adaptation with speakers which are different from the test base. This demonstrates that the voices of ageing people have common characteristics. Nevertheless, a little part of senior people (G03 group) can be characterized with poorer results of speech recognition. This group is not composed of the oldest people.

Group	Gender M/F	Age min-max	WER_1 (%)	WER_2 (%)
G01	4/13	70-92	13.58	5.54
G02	3/12	63-94	33.32	14.41
G03	4/0	62-84	65.38	29.97
All	11/25	62-94	27.56	11.95

Table 3. Comparison of WER with the generic acoustic model (WER_1) and the MLLR adapted model (WER_2) for the 36 *Aged Speech* group

CONCLUSION AND PERSPECTIVES

In this paper we present two studies aiming at investigating the feasibility of speech-based interactive systems. In the first study, the acceptability of a voice interface as part of the smart home was investigated. Voice interfaces appear to be better accepted by the seniors than more intrusive solutions such as video cameras. Otherwise, the "key-word" form for commands is highly accepted rather than the sentence based command. An interesting finding that came up is their overall acceptance provided the system does not drive them to a lazy lifestyle by taking control of everything. Smart homes could give seniors more ability to control their daily living.

The second study is related to the adaptation of speech recognition technologies to the senior population. Therefore, we recorded two specific speech corpora (Voice-Age and ERES38) which were analyzed in a semi automatic manner to reveal the aged-voice characteristics. Some phonemes are more affected by age than others, nasal vowels and

consonants. Moreover, the absence of voicing is the main factor of degradation. Our current work is to complete this corpus in order to obtain more generic senior models. The CIRDO system[27] will take advantage of theses models to integrate new services for autonomy increase and make easier the support for relatives and caregivers.

Acknowledgments

These studies were funded by the French National Agency (project Sweet-Home[26] ANR-2009-VERS-011 and project CIRDO ANR-2010-TECS-012). The authors would like to thank the participants (seniors, relatives and caregivers) who accepted to perform the experiment and agreed to participate to the speech recordings.

References

1. F. Portet, M. Vacher, C. Golanski, C. Roux, B. Meillon, Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects, *Personal and Ubiquitous Computing*, http://dx.doi.org/10.1007/s00779-011-0470-5, pages 1-18 (in press).
2. Chan, M., Estève, D., Escriba, C., and Campo, E., A review of smart homes-Present state and future challenges, *Computer Methods and Programs in Biomedicine*, 91(1):55–81, 2008.
3. Callejas, Z. and López-Cózar, R., Designing smart home interfaces for the elderly, *SIGACCESS Newsletter*, 95, 2009.
4. V. Rialle, C. Ollivet, C. Guigui, C. Hervé, What do family caregivers of alzheimer's disease patients desire in smart home technologies? Contrasted results of a wide survey, *Methods of Information in Medicine,* 47(1):63–69, 2008.
5. Vacher, M., Portet, F., Fleury, A., and Noury, N., Development of audio sensing technology for ambient assisted living: Applications and challenges, *International Journal of E-Health and Medical Communications*, 2(1):35–54, 2011.
6. L.R. Rabiner, A tutorial on hidden Markov models ans selected applications in speech recognition, *Proceedings of the IEEE*, 77(2):257-286, 1989.
7. G. Cornet, A. Franco, V. Rialle, P. Rumeau, chap. Les gérontechnologies au cœur de l'innovation hospitalière et médico-sociale, *Société Française de Technologie pour l'Autonomie et de Gérontechnologie*, 703 :53-58, 2007.
8. M. Gerosa, D. Giuliani, F.Brugnara, Towards age-independent acoustic modelling, *Speech Communication*, 51(6):499-509, 2009.
9. A. Sharkey, N. Sharkey, Granny and the robots: ethical issues in robot care for the elderly, *Ethics ans Information Technology*, pages 1-14, in press.
10. J. van Hoof, H.S.M. Kort, P. Markopoulos, M. Soede, Ambiant intelligence, ethics and privacy, *Gerontechnology*, 6(3):155-163, 2007.
11. W. Ryan, K. Burk, Perceptual and acoustic correlates in the speech of males, *Journal of Communication Disorders*, 7:181-192, 1974.
12. N. Taked, G. Thomas, C. Ludlow, Aging effects on motors units in the human thyroarytenoid muscle, *Laryngoscope*, 110:1018-1025, 2000.
13. P. Mueller, R. Sweeney, L. Baribeau, Acoustic and morphologic study of the senescent voice, *Ear, Nose and Throat Journal*, 63:71-75, 1984.
14. A. Baba, S. Yoshizawa, M. Yamada, A. Lee, K. Shikano, Acoustic models of the elderly for large-vocabulary continuous speech recognition, *Electronics and Communications in Japan*, 87(2):49-57, 2004.
15. R. Vipperla, S. Renals, J. Frankel, Longitudinal study of ASR performance on ageing voices, in Proc. Interspeech, pages 2550-2553, 2008.
16. R. Privat, N. Vigouroux, P. Truillet, Etude de l'effet du vieillissement sur les productions langagières et sur les performances en reconnaissance automatique de la parole, *Revue Parole*, 31-32 :281-318, 2004.
17. M. Vacher, A. Fleury, J.-F. Serignat, N. Noury, H. Glasson, Preliminary evaluation of speech/sound recognition for telemedicine application in a real environment, in Proc. InterSpeech, 1:496-499, 2008.
18. F. Aman, M. Vacher, S. Rossato, R. Dugheanu, F. Portet, J. Legrand, Y. Sasa, Performance des modèles acoustiques pour des voix de personnes âgées en vue de l'adaptation des systèmes de RAP, Journées d'Etudes sur la Parole, pages 1-8, 2012.
19. V. Taler, N. Phillips, Language performance un Alzheimer's disease and mild cognitive impairment : A comparative review, *Journal of Clinical and Experimental Neurophysiology*, 30(5) :501-556, 2008.
20. http://www.speech.cs.cmu.edu/, Speech at CMU.
21. L. Lamel, J. Gauvain, M. Eskenazi, BREF, a large vocabulary spoken corpus for French, in Proc. EUROSPEECH 91, Vol. 2, pages 505-508, 1991.
22. J. Rodin, Aging and health: effects of the sense of control, *Science*, 233(4770):1271–1276, 1986.
23. M. Hamill, V. Young, J. Boger, A. Mihailidis, Development of an automated speech recognition interface for personal emergency response systems, *Journal of NeuroEngineering and Rehabilitation*, 6:1-11 2009.
24. T. Koskela, K. Väänänen-Vainio-Mattila, Evolution towards smart home environments: empirical evaluation of three user interfaces, *Personal and Ubiquitous Computing*, 8:234–240, 2004.
25. M. Rantz, R. Porter, D. Cheshier, D. Otto, C. Servey, R. Johnson, M. Aud, M. Skubic, H. Tyrer, Z. He, Z., G. Demiris, G. Alexander, G. Taylor, TigerPlace, a State-Academic-Private project to revolutionize traditional Long-Term care, *Journal of Housing For the Elderly*, 22(1):66-85, 2008.
26. http://sweet-home.imag.fr/, The Sweet-Home project.
27. http://liris.cnrs.fr/cirdo/, The CIRDO project.