Development of Automatic Speech Recognition Techniques for Elderly Home Support: Applications and Challenges

Michel Vacher, Frédéric Aman, Solange Rossato, François Portet

HAL Id: hal-01179246
https://hal.archives-ouvertes.fr/hal-01179246

Submitted on 22 Jul 2015

Development of automatic speech recognition techniques for elderly home support: Applications and Challenges

Michel Vacher[1], Frédéric Aman[1], Solange Rossato[2], and François Portet[2]

[1] CNRS, LIG, F-38000 Grenoble, France,
Michel.Vacher@imag.fr - Frederic.Aman@imag.fr,
[2] Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France,
Solange.Rossato@imag.fr - Francois.Portet@imag.fr,
Laboratoire d'Informatique de Grenoble, GETALP
WWW home page: http://www.liglab.fr/util/annuaire

Abstract. Vocal command may have considerable advantages in terms of usability in the AAL domain. However, efficient audio analysis in smart home environment is a challenging task in large part because of bad speech recognition results in the case of elderly people. Dedicated speech corpora were recorded and employed to adapted generic speech recognizers to this type of population. Evaluation results of a first experiment allowed to draw conclusions about the distress call detection. A second experiments involved participants who played fall scenarios in a realistic smart home, 67% of the distress calls were detected online. These results show the difficulty of the task and serve as basis to discuss the stakes and the challenges of this promising technology for AAL.

Keywords: Automatic Speech Recognition, Aged Voices, Home Automation, Vocal Command, Distress Call, Ambient Assisted Living

1 Introduction

Life expectancy has increased in all countries of the European Union in the last decade. In the beginning of 2013, 9% of the people in France were at least 75 years old. The number of dependent elderly people will increase by 50% by 2040 according to INSEE institute [12]. The notion of dependency is based on the alteration of physical, sensory and cognitive functions having as a consequence the restriction of the activities of daily living, and the need for help or assistance of someone for regular elementary activities [7]. While the transfer of dependant people to nursing homes has been the *de facto* solution, a survey shows that 80% of people above 65 years old would prefer to stay living at home if they lose autonomy [10].

The aim of Ambient Assisted Living (AAL) is to compensate the alteration of physical, sensory and cognitive functions, that are cause of activity restrictions, by technical assistance or environmental management through the use of

Information and Communication Technology (ICT) [3] as well as to anticipate and respond to the need of persons with loss of autonomy while AAL solutions are being developed in robotics, home automation, cognitive science, computer network, etc.

We will focus on the domain of smart homes [6][11][25] which are a promising way to help elderly people to live independently. In the context of AAL, the primary tasks of the smart homes are the followings:

- *to support disabled users* via specialized devices (rehabilitation robotics, companion robot, wheelchair, audio interface, tactile screen, etc.);
- *to monitor the users* in their own environment at home thanks to home automation sensors or wearable devices (accelerometer or physiological sensors recording heart rate, temperature, blood pressure, glucose, etc.);
- *to deliver therapy* thanks to therapeutic devices;
- *to ensure comfort and reassurance* thanks to intelligent household devices, smart objects and home automation.

It is worth noting that, within this particular framework, intelligent house equipment (e.g., motion sensors) and smart leisure equipment (interactive communication systems and intelligent environmental control equipments) are particularly useful in case of emergency to help the user to call his relatives, as well as transmitting automatically an alert when the user is not able to act himself. At this time, an other research domain related to energy efficiency is emerging.

Techniques based on very simple and low cost sensors (PIR)[13] or on video analysis[19] are very popular, however they can not be used for interaction purpose unless they are completed by a tactile device (smart phone), while Vocal-User Interface (VUI) may be well adapted because a natural language interaction is relatively simple to use and well adapted to people with reduced mobility or visual impairment [27]. However, there are still important challenges to overcome before implementing VUIs in a smart home [36] and this new technology must be validated in real conditions with potential users [25].

A rising number of recent projects in the smart home domain include the use of Automatic Speech Recognition (ASR) in their design [4][5][26][15][16][9][22] and some of them take into account the challenge of Distant Speech Recognition [34][23]. These conditions are more challenging because of ambient noise, reverberation, distortion and acoustical environment influence. However, one of the main challenges to overcome for successful integration of VUIs is the adaptation of the system to elderly. From an anatomical point of view, some studies have shown age-related degeneration with atrophy of vocal cords, calcification of laryngeal cartilages, and changes in muscles of larynx [32][24]. Thus, ageing voice is characterized by some specific features such as imprecise production of consonants, tremors and slower articulation [29]. Some authors [1][37] have reported that classical ASR systems exhibit poor performances with elderly voice. These few studies were relevant for their comparison between ageing voice vs. non-ageing voice on ASR performance, but their fields were quite far from our

---

[3] http://www.aal-europe.eu/

Table 1. Speech recognition technologies in smart homes for assistive technologies

Project or study	Users	Characteristics of the voice	Conditions of record	Studio or smart home	Vocabulary
S1 [21]	Adults	Neutral: vocal command (wheelchair)	Close	Boston Home	Large
S2 [26]	Elderly	Neutral: ageing well	Close	MLDC corpus	Large
S3 [38]	Elderly	Professional: defense speech	Close	Supreme Court (US)	Large
P1 [15]	Dysarthric	Realistic: vocal command	Close	Studio	Short and adaptative
P2 [23]	Adults	Neutral: vocal command	Distant	SmH	Large
P3 [5]	-	Neutral: Dialog companion	Close	-	Large
P4 [22]	Adults	Neutral: control of appliances/devices	Distant	DIRHA home	Large
P5 [9]	Dysarthric	Neutral: vocal command (assistive technologies)	Microphone array	-	Short
P6 [16]	Adults	Neutral: alarm confirmation	Micro. phone	-	"yes"/"no"
P7 [4]	Dysarthric	Realistic: environmental control	Close	UA-speech corpus	Short
P8 [34]	Adults/ elderly	Neutral: home automation	Distant	DOMUS	Medium
P9 (This study)	Elderly	Neutral/expressive call for help	Short	Medical institutions	Short

topic of automation commands recognition, and no study was done in French language, except for pathologic voices [14].

In this paper, we present the results of our study related to a system able to detect the call of elderly for emergency when they are in a distress case.

2 State of the Art

A large number of research projects were related to assistive technologies, among them House_n [18], Casas [8], ISpace [17], Aging in Place [31], DesdHIS [13], Ger'Home [41] or Soprano [39]. A great variety of sensors were used like wearable video cameras, embedded sensors, medical sensors, switches and infrared detectors. The main trends of these projects were related to activity recognition, health status monitoring and cognitive stimulation. Thanks to recent advances in microelectronics and ICT, smart home equipments could operate efficiently with low energy consumption and could be available at low prices.

Table 2. Studies and projects related to speech recognition of aged people

Proj.	Principal author or acronym	Ref.	Proj.	Acronym	Ref.
S1	Li, W.	[21]	P4	DIRHA	[22]
S2	Pellegrini, T.	[26]	P5	homeSERVICE	[9]
S3	Vipperla, R.	[38]	P6	PERS	[16]
P1	ALADIN	[15]	P7	PIPIN	[4]
P2	COMPANIONABLE	[23]	P8	SWEET-HOME	[34]
P3	COMPANIONS	[5]	**P9**	**CIRDO**	**(This study)**

Regarding speech technologies, the corresponding studies and projects are in most cases related to smart homes and assistive technologies. Table 1 summaries their principal characteristics[4]. Among these projects, COMPANIONABLE, COMPANIONS and DIRHA, while aiming at assisting elderly people, mostly performed studies including typical non-aged adults; the greatest of the SWEET-HOME studies were related to adult voices but some aged and visually impaired people took part in one experiment. Automatic recognition of elderly speech was mainly studied for English by Vipperla et al. [38] and for Portuguese by Pellegrini et al. [26]. These two studies confirmed that the performances of standard recognizers decrease in the case of aged speakers. Vipperla et al. used the SCOTUS speech corpus which is the collection of the audio recordings of the proceedings of the Supreme court of the United States of America. This corpus allowed them to analyze the voice of a same speaker over more than one decade. By contrast, ALADIN, homeSERVICE, and PIPIN considered the case of Alzheimer's voices, which is a more difficult task than for typical voice because of the cognitive and perceptual decline affecting this part of the population since it may impact the grammatical pronunciation and flow of speech which current speech recognizers can not handle.

Figure 1 describes the general organisation of an Automatic Speech Recognition systems (ASR), the decoder is in charge of phone retrieval in a sequence of feature vectors extracted from the sound, the simplest and more commonly used are Mel-Frequency Cepstral Coefficients (MFCCs) [40]. Phones are the basic sound units and are mostly represented by a continuous density Hidden Markov Model (HMM). The decoder tries to find the sequence of words $\widehat{W}$ that match the input signal $\mathbf{Y}$:

$$\widehat{W} = \arg \max_W \left[p(\mathbf{Y}|W)\, p(W) \right] \tag{1}$$

The likelihood $p(\mathbf{Y}|W)$ is determined by an Acoustic Model (AM) and the prior $p(W)$ by a Language Model (LM). ALADIN is based on principles radically different from those of classical ASRs and uses a direct decoding thanks to Non-negative Matrix Factorization (NMF) and does not use any AM or LM.

ASRs have reached good performances with close talking microphones (e.g. head-set), but the performances decrease significantly as soon as the microphone

---

[4] The correspondence between project number and reference is given in Table 2
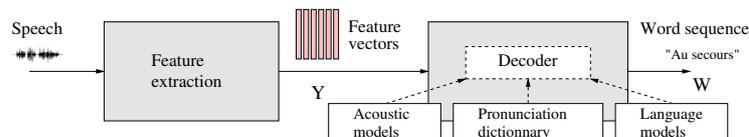
Fig. 1. Architecture of a classical ASR

is moved away from the mouth of the speaker (e.g., when the microphone is set in the ceiling). This deterioration is due to a broad variety of effects including reverberation and presence of undetermined background noise. Distant speech recognition is the major aim of DIRHA, COMPANIONABLE and SWEET-HOME. The SWEET-HOME project aimed at controlling an intelligent home automation system by vocal command, and a study done in this framework showed that good performances can be obtained thanks to Acoustic Models trained on the same conditions as the target model and using multiple channels [35].

Studies in the Natural Language Processing (NLP) domain require the use of corpora which are essential at all steps of the investigations and particularly during the model training and the evaluation. To the best of our knowledge, very few corpora are related to ageing voices in French [14]. The different available corpora are stems of projects related to the study of French language like the "Corpus de Français Parlé Parisien des années 2000[5]". This corpus is made of recordings of inhabitants of different districts of Paris in order to study the influence of French spoken language over France and the French speaking world. The "Projet Phonologie du Français Contemporain[6]" is a database of records according to the region or the country. The records of 38 elderly people (above 70 years old) are included, each record is made of a word list, a small text and two interviews. Other available sources come from videos of testimonies of Shoah survivals and recorded in the framework of "Mémorial de la Shoah[7]" which collect testimonies and organize conferences. These videos are not annotated. This corpus is then a collection of interviews and spontaneous speech.

As no study was done with the purpose of facilitating the communication and the detection of distress calls and given that no corresponding corpus exists in French, the first challenge was to record speech corpora uttered by aged people in order to study the characteristics of their voices and explore ways to adapt ASR systems in order to improve their performances for this population category. The second challenge was related to the evaluation of the usability and the acceptance of systems based on speech recognition by their potential users in a smart home.

---

[5] http://ed268.univ-paris3.fr/syled/ressources/Corpus-Parole-Paris-PIII
[6] http://www.projet-pfc.net/
[7] http://www.memorialdelashoah.org/

3 Corpus acquisition and analysis system

Therefore, in a first step, we recorded two corpora AD80 and ERES38 adapted to our application domain. ERES38 was used to adapt the acoustic models of a standard ASR and we evaluated the recognition performances on the AD80 corpus. Moreover, we drawed some conclusions about the performance differences of ASR between non-aged and elderly speakers.

The first corpus ERES38 was recorded by 24 elderly people (age: 68-98 years) in French nursing homes. It is made of text reading (48min) and interviews (4h 53min). This corpus was used for acoustic model adaptation.

The second corpus AD80 was recorded by 52 non-aged speakers (age: 18-64 years) in our laboratory and by 43 elderly people (62-94 years) in medical institutions. This corpus is made of text readings (1h 12min) and 14,267 short sentences (4h 49min). There are 3 types of sentences: -*distress calls* ("I fell"), -*home automation commands* ("switch the light on") and -*casual* ("I drink my coffee"). The distress calls are the sentences that a person could utter during a distress situation to request for assistance, for example after he fell. The determination of a list of these calls is a challenging task. Our list was defined in collaboration with the GRePS laboratory after a bibliographical study[2] and in the prolongation of previous studies [36].

This corpus was used firstly for ASR performance comparison between the two groups (aged/non-aged) and in a second step to determine if acoustic model adaptation could allow the detection of distress or call for help sentences. It was necessary to assess the level of loss of functional autonomy of the 43 elderly speakers. Therefore, a GIR [30] score was obtained after clinicians filled the AGGIR grid (French national test) to classify the person in one of the six groups: GIR 1 (total dependence) to GIR 6 (total autonomy).

The last corpus is the Cirdo-set corpus [3]. This corpus was recorded in the Living Lab of the LIG laboratory by 13 young adults (32min 01s) and 4 elderly people (age: 61-83 years, 28min 54s) which played 4 scenarios relative to *fall*, one to *blocked hip* and two True Negative (TN) scenarios. These scenarios included calls for help which are identical to some of the corresponding sentences of AD80. The audio records of the Cirdo-set corpus were then used for evaluation purpose of call for help detection in realistic conditions. These are full records, therefore the speech events have to be extracted thanks to an online analysis system. This process will be presented in Section 4. Moreover, the recording microphone was set in the ceiling and not as usual at a short distance in front of the speaker but in Distant Speech conditions.

The corpora were processed by the ASR of the CMU toolkit Sphinx3 [20]. The acoustic vectors are composed of 13 MFCC coefficients, their first and second derivatives. The Acoustic Model (AM) is context-dependent with 3-state left-to-right HMM. We used a generic AM trained with BREF120, a corpus made of 100 hours of French speech. The language model was a 3-gram-type LM resulting from the combination of a generic language model (with a 10% weight) and the domain one (with 90% weight). The generic LM resulting from French news collected in the Gigaword corpus was 1-gram with 11,018 words. The domain

Table 3. WER using the generic acoustic model AM

Subject	Non-elderly	Elderly	Difference
Women	10.4% (27 speakers)	40.3% (32 speakers)	+29.9%
Men	11.7% (25 speakers)	61.3% (11 speakers)	+49.6%
Average	11.0%	45.7%	34.7%
Standard deviation	6.4%	16.8%	10.4%

LM trained from the AD80 corpus was composed of 88 1-gram, 193 2-gram and 223 3-gram.

The target is that only the sentences of interest could be recognized by the system (i.e., not when they are receiving a phone call from their relatives) [27]. Therefore, only two categories of the sentences are relevant to the system and must be taken into consideration: home automation commands and calls related to a distress situation. The other sentences must be discarded and it is therefore necessary to determine whether the resulting output from the ASR is part of one of the two categories of interest thanks to a measure distance. This measure is based on a Levenshtein distance between each output and typical sentences of interest. In this way, casual sentences are excluded.

4 Adaptation of the system to elderly voices and detection of distress calls

To assess ASR performances, the most common measure is the Word Error Rate (WER) which is defined as follows:

$$WER = \frac{S + D + I}{N} \qquad (2)$$

S is the number of substitutions, D the number of deletions, I the number of insertions and N the number of words in the reference. As shown in Table 3, when performing ASR using the generic acoustic model on the distress/home automation sentences of the AD80 corpus, we obtained an average WER of 45.7% for the elderly group in comparison with an average WER of 11% for the non-elderly group. These results indicate a significant decrease in performance for elderly speech and we can notice an important scattering of the results for this kind of voice as well as a higher recognition rate for women as supported by the state of the art. It is thus clear that the generic AM is not adapted to the elderly population and then specific models must be used.

Thanks to a Maximum Likelihood Linear Regression (MLLR), the text readings of the ERES38 corpus were used to obtain 3 specific aged AMs from the generic AM: AM_G (men and women), AM_W (women) and AM_M (men). Table 4 gives the obtained results and indicates a significant improvement of the performances. An ANOVA analysis allowed us to conclude that: (1) there is no significant difference between generic and specific models for non-aged speakers;

Table 4. WER using the specific acoustic models (***: $p < 0.001$)

Subject	AM (generic)	AM_G (global)	AM_W (women)	AM_M (men)
Non-aged women	10.4%	7.6%	7.3%	-
Non-aged men	11.7%	11.1%	-	11.8%
Elderly women	40.3%	14.6%***	15.4%***	-
Elderly men	61.3%	25.7%***	-	22.3%***

(2) the difference between generic and specific models is significant; (3) there is no significant difference between the specific models (AM_G, AM_W, AM_M) and thus the use of a unique global model is possible. In the case of aged speaker, the dispersion of the performances is very important whatever acoustic model is chosen (e.g., $WER_{AM\_G} = 17.4\%$ and $\sigma_{AM\_G} = 10.3\%$). This dispersion is due to bad performances encountered with some speakers, they are those who suffer of an important loss of functional autonomy (GIR 2 or 3) and then are less likely to live alone at their own home.

As reported in Section 3, only sentences related to a call for help or home automation management have to be analysed, the other one (i.e., casual) being rejected. Every sentence whose distance to the distress category was above a threshold th was rejected.

For our study, we considered the sentences of AD80 uttered by elderly speakers, namely 2,663 distress sentences, 434 calls for caregivers and 3,006 casual sentences. The ASR used AM_G as model. The threshold th of the filter was chosen in such way that the sensibility Se and the specificity Sp were equal ($th = 0.75$, $Se = Sp = 85.7\%$). It should be noted that, due to the WER, 4% of the selected sentences were put in the correct category but did not correspond to the sentence as it was pronounced. Regarding the distress sentences and calls to caregivers, 18% were selected with confusion. Consequently, the main uncertainty concerns above all the way in which the call must be treated.

5 Evaluation of the detection in real conditions with the audio components of the Cirdo-set corpus

For the evaluation of the detection of distress calls *in situ*, we used the Cirdo-set corpus which was recorded in a Living Lab. In order to extract the sentences pronounced by the speakers during the scenarios, we used CirdoX, an online audio analyser in charge of detecting the audio events and discriminating between noise and speech. The diagram of CirdoX is presented Figure 2. CirdoX is able to capture signal from microphones or to analyse previous audio records on 8 channels, we used it in a mono-channel configuration. The detection of each audio event is operated online thanks to an adaptive threshold on the high level components of the wavelet transform of the input signal. Each audio event is then classified into speech or noise. The GMM classifier was trained with the Sweet-Home corpus [33] recorded in a smart home. The ASR was Sphinx3 as mentioned above.
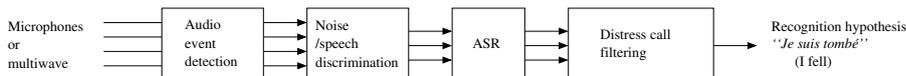
Fig. 2. Architecture of the CirdoX online analyser

CirdoX detected 1950 audio events including 322 speech events, 277 of them were calls for help. 204 were analysed as speech and 73 as noise mainly due to a strong presence of environmental noise at the moment of the record. Because of the distant speech conditions, the acoustic model was adapted with sentences of the Sweet-Home corpus recorded in similar conditions [33]. Regarding the calls for help sent to the ASR, the WER was 49.5% and 67% of the calls were detected. These results are far from perfect but they were obtained under harsh conditions. Indeed, the participants played scenarios which included falls on the floor and the participants generated a lot of noise sounds which were often mixed with speech. Therefore, the performances would have been better if the call were uttered after the fall.

Moreover, these results were obtained using a classical ASR as Sphinx but significant improvements were made recently in speech recognition and incorporated in the KALDI toolkit [28]. Off line experiments were done in this framework on the "Interaction Subset" of the Sweet-Home corpus [35]. This corpus is made of records in a smart home equipped with a home automation system including more than 150 sensors and actuators. The home automation network is driven by an Intelligent Controller able to take a context aware decision when a vocal command is recognised. Among other things, the controller must choose what room and what lamp are concerned. The corresponding sentences are home automation vocal commands pronounced by participants who played scenarios of the everyday life. They asked for example to switch on the light or to close the curtains while they are eating breakfast or doing the dishes.

The speech events, for instance 550 sentences (2559 words) including 250 orders, questions and distress calls (937 words), were extracted using PATSH, an online audio analyser which is similar to CirdoX. The original ASR performance with a decoding on only one channel was WER=43.2%, DER=41% [34], DER being defined as the Detection Error Rate of the home automation commands. Thanks to 2 more sophisticated adaptation techniques, namely Subspace GMM Acoustic Modelling SGMM) and feature space MLLR (fMLLR) significant improvement were brought which led to WER=49%, DER=13.6%. The most important contribution to the DER was due to missed speech utterances at the detection or speech/sound discrimination level. This significant improvement from the experimental condition was obtained in off line conditions and the most important effort must be related to adapt and integrate these new techniques in an online audio analyser, i.e. CirdoX.

6   Conclusion

Regarding the technical aspect, our study showed first of all that thanks to the record of a short corpus by elderly speakers (ERES38, 48min), it is possible to adapt the acoustic models (AM) of a generic ASR and to obtain recognition performances in the case of elderly voices close to those of non-aged speakers (WER about 10% or 15%), except for elderly affected by an important level of loss of functional autonomy. Therefore the detection of distress sentences is efficient and the sensibility is 85%. Our experiment involving the Cirdo-set corpus recorded in *in-situ* conditions gave lower results due to the harsh conditions, the participants falling as they called for help and only 67% of the calls were detected. However new adaptation techniques may improve significantly the results as soon as they will be integrated in an online audio analyser.

People who participated to the experiments were excited and wanted to use such a technology in their own environment, as it was reported in some studies [27]. However, the use of a short vocabulary is necessary in order to obtain good performances, so an important difficulty is related to the difficulty of defining which sentences would be pronounced during a fall or a distress situation. Thanks to the collaboration with the GRePS laboratory some of those were incorporated in the AD80 corpus but it is not sufficient for a real application. There is no adequate corpus and the potential users exhibit great difficulties in remembering the sentences they pronounced in such situations. Therefore an important effort will consist in the necessary adaptation of the language models (ML) to the user in the long life term.

References

1. Baba, A., Lee, A., Saruwatari, H., Shikano, K.: Speech recognition by reverberation adapted acoustic model. In: ASJ General Meeting. pp. 27–28 (2002)
2. Bobillier Chaumon, M., Bekkadja, S., Cros, F., Cuvillier, B.: The user-centered design of an ambient technology for preventing falls at home. Gerontechnology 13(2), 169 (2014)
3. Bouakaz, S., Vacher, M., Bobillier-Chaumon, M.E., Aman, F., Bekkadja, S., Portet, F., Guillou, E., Rossato, S., Desserée, E., Traineau, P., Vimon, J.P., Chevalier, T.: CIRDO: Smart companion for helping elderly to live at home for longer. Innovation and Research in BioMedical engineering (IRBM) 35(2), 101–108 (Mar 2014)
4. Casanueva, I., Christensen, H., Hain, T., Green, P.: Adaptive speech recognition and dialogue management for users with speech disorders. In: Interspeech 2014. pp. 1033–1037 (2014)
5. Cavazza, M., de la Camara, R.S., Turunen, M.: How was your day? A companion ECA prototype. In: AAMAS. pp. 1629–1630 (2010)

6. Chan, M., Estève, D., Escriba, C., Campo, E.: A review of smart homes– Present state and future challenges. Computer Methods and Programs in Biomedicine 91(1), 55–81 (2008)
7. Charpin, J.M., Tlili, C.: Perspectives démographique et financières de la dépendance, Rapport du groupe de travail sur la prise en charge de la dépendance. Tech. rep., Ministère des solidarités et de la cohésion sociale, 60 p., Paris (2011)
8. Chen, C., Cook, D.J.: Behavior-based home energy prediction. In: IEEE Intelligent Environments. pp. 57–63 (2012)
9. Christensen, H., Casanueva, I., Cunningham, S., Green, P., Hain, T.: homeService: Voice-enabled assistive technology in the home using cloud-based automatic speech recognition. In: 4th Workshop on Speech and Language Processing for Assistive Technologies. pp. 29–34 (2013)
10. CSA: Les français et la dépendance. http://www.csa.eu/fr/s26/nos-sondages-publies.aspx (2003), accessed: 12/03/2013
11. De Silva, L., Morikawa, C., Petra, I.: State of the art of smart homes. Engineering Applications of Artificial Intelligence 25(7), 1313–1321 (2012)
12. Duée, M., Rebillard, C.: La dépendance des personnes âgées : une projection en 2040. Données sociales - La société française pp. 613–619 (2006)
13. Fleury, A., Vacher, M., Noury, N.: SVM-based multi-modal classification of activities of daily living in health smart homes: Sensors, algorithms and first experimental results. IEEE TITB 14(2), 274 –283 (2010)
14. Gayraud, F., Lee, H., Barkat-Defradas, M.: Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. Clinical linguistics & phonetics 25(3), 198–209 (2011)
15. Gemmeke, J.F., Ons, B., Tessema, N., Van Hamme, H., Van De Loo, J., De Pauw, G., Daelemans, W., Huyghe, J., Derboven, J., Vuegen, L., Van Den Broeck, B., Karsmakers, P., Vanrumste, B.: Self-taught assistive vocal interfaces: an overview of the ALADIN project. In: Interspeech 2013. pp. 2039–2043 (2013)
16. Hamill, M., Young, V., Boger, J., Mihailidis, A.: Development of an automated speech recognition interface for personal emergency response systems. Journal of NeuroEngineering and Rehabilitation 6(1), 1–26 (2009)
17. Holmes, A., Duman, H., Pounds-Cornish, A.: The iDorm: gateway to heterogeneous networking environments. In: International ITEA workshop on Virtual Home Environments. pp. 20–37 (2002)
18. Intille, S.S.: Designing a home of the future. IEEE Pervasive Computing 1(2), 76–82 (2002)
19. König, A., Crispim, C., Derreumaux, A., Bensadoun, G., Petit, P.D., Bremond, F., David, R., Verhey, F., Aalten, P., Robert, P.: Validation of an automatic video monitoring system for the detection of instrumental activities of daily living in dementia patients. Journal of Alzheimer's Disease 44(2), 675–685 (2015)
20. Lee, K.F., Hon, H.W., Reddy, R.: An overview of the SPHINX speech recognition system. IEEE TASSP 38(1), 35–45 (1990)
21. Li, W., Glass, J., Roy, N., Teller, S.: Probabilistic dialogue modeling for speech-enabled assistive technology. In: SLPAT 2013. pp. 67–72 (2013)
22. Matassoni, M., Astudillo, R.F., Katsamanis, A., Ravanelli, M.: The DIRHA-GRID corpus: baseline and tools for multi-room distant speech recognition using distributed microphones. In: Interspeech 2014. pp. 1613–1617 (2014)
23. Milhorat, P., Istrate, D., Boudy, J., Chollet, G.: Hands-free speech-sound interactions at home. In: EUSIPCO 2012. pp. 1678 –1682 (aug 2012)
24. Mueller, P., Sweeney, R., Baribeau, L.: Acoustic and morphologic study of the senescent voice. Ear, Nose, and Throat Journal 63, 71–75 (1984)

12

25. Peetoom, K.K.B., Lexis, M.A.S., Joore, M., Dirksen, C.D., De Witte, L.P.: Literature review on monitoring technologies and their outcomes in independently living elderly people. Disability and Rehabilitation: Assistive Technology pp. 1–24 (2014)

26. Pellegrini, T., Trancoso, I., Hämäläinen, A., Calado, A., Dias, M.S., Braga, D.: Impact of Age in ASR for the Elderly: Preliminary Experiments in European Portuguese. In: Proceedings of Advances in Speech and Language Technologies for Iberian Languages (IberSPEECH 2012 Conference). pp. 139–147 (2012)

27. Portet, F., Vacher, M., Golanski, C., Roux, C., Meillon, B.: Design and evaluation of a smart home voice interface for the elderly – Acceptability and objection aspects. Personal and Ubiquitous Computing 17(1), 127–144 (2013)

28. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The Kaldi Speech Recognition Toolkit. In: ASRU 2011 (2011)

29. Ryan, W., Burk, K.: Perceptual and acoustic correlates in the speech of males. Journal of Communication Disorders 7, 181–192 (1974)

30. Site officiel de l'administration française: Allocation personnalisée d'autonomie (Apa) : grille Aggir. http://vosdroits.service-public.fr/F1229.xhtml

31. Skubic, M., Alexander, G., Popescu, M., Rantz, M., Keller, J.: A smart home application to eldercare: Current status and lessons learned. Technology and Health Care 17(3), 183–201 (2009)

32. Takeda, N., Thomas, G., Ludlow, C.: Aging effects on motor units in the human thyroarytenoid muscle. Laryngoscope 110, 1018–1025 (2000)

33. Vacher, M., Lecouteux, B., Chahuara, P., Portet, F., Meillon, B., Bonnefond, N.: The Sweet-Home speech and multimodal corpus for home automation interaction. In: The 9th edition of the Language Resources and Evaluation Conference (LREC). pp. 4499–4506. Reykjavik, Iceland (2014)

34. Vacher, M., Lecouteux, B., Istrate, D., Joubert, T., Portet, F., Sehili, M., Chahuara, P.: Experimental Evaluation of Speech Recognition Technologies for Voice-based Home Automation Control in a Smart Home. In: SLPAT 2013. pp. 99–105 (2013)

35. Vacher, M., Lecouteux, B., Portet, F.: Multichannel Automatic Recognition of Voice Command in a Multi-Room Smart Home : an Experiment involving Seniors and Users with Visual Impairment. In: Interspeech 2014. pp. 1008–1012 (Sep 2014)

36. Vacher, M., Portet, F., Fleury, A., 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 (Mar 2011)

37. Vipperla, R.C., Wolters, M., Georgila, K., Renals, S.: Speech input from older users in smart environments: Challenges and perspectives. In: HCI International: Universal Access in Human-Computer Interaction Environments. pp. 117–126 (2009)

38. Vipperla, R., Renals, S., Frankel, J.: Longitudinal study of ASR performance on ageing voices. In: Interspeech 2008. pp. 2550–2553 (2008)

39. Wolf, P., Schmidt, A., Klein, M.: SOPRANO - An extensible, open AAL platform for elderly people based on semantic contracts. In: 3rd Workshop on Artificial Intelligence Techniques for Ambient Intelligence (ECAI 08) (2008)

40. Young, S.: Handbook of Speech Processing, chap. HMMs and Related speech Recognition Technologies, pp. 539–557. Springer (2008)

41. Zouba, N., Bremond, F., Thonnat, M., Anfosso, A., Pascual, E., Mallea, P., Mailland, V., Guerin, O.: A computer system to monitor older adults at home: Preliminary results. Gerontechnology Journal 8(3), 129–139 (July 2009)