Analysing the Performance of Automatic Speech Recognition for Ageing Voice: Does it Correlate with Dependency Level?

Frédéric Aman, Michel Vacher, Solange Rossato, François Portet

Analysing the Performance of Automatic Speech Recognition for Ageing Voice: Does it Correlate with Dependency Level?

Frédéric Aman, Michel Vacher, Solange Rossato and François Portet

LIG, UMR5217 UJF/CNRS/Grenoble-INP/UPMF, 38041 Grenoble, France

{frederic.Aman, Michel.Vacher, Solange.Rossato, Francois.Portet}@imag.fr

Abstract

Ambient Assisted Living aims at providing assistance by allowing people with special needs to perform tasks which they have increasing difficulty with and to provide reassurance through surveillance in order to detect distress and accidental falls. Aged people are among the ones who might benefit from advances in ICT to live as long as possible in their own home. Voice-base smart home is a promising way to provide AAL, but even mature technologies must be evaluated from the perspective of its potential beneficiaries. In this paper, we investigate which characteristics of the ageing voice that challenge a state of the art ASR system. Though in the literature, chronological age is retained as the sole factor predicting decrease in performance, we show that degree of loss of autonomy is even more correlated to ASR performance.

Index Terms: Ambient Assisted Living (AAL), Dependency, Elderly speech, Voice command

1. INTRODUCTION

With advances in medicine, life expectancy has increased. However, this phenomenon coupled with a low birthrate has led to an ageing population in industrialised countries. To help elderly people to live as long as possible in their home, solutions have been developed based on robotics, automation, cognitive science, and computer networks. These solutions are being developed to compensate their possible physical or mental decline to keep them with a good degree of autonomy. The aim is to provide assistance by allowing them to perform tasks which they have increasing difficulty with and to provide reassurance through surveillance in order to detect distress and accidental falls. Such a system must allow independence of elderly while facilitating social contact, with a major impact on well-being and health. In addition, it helps caregivers and reassure relatives. However, technological solutions must be able to adapt to the needs and the specific capacities of this population. Indeed, elderly are often confused by complex interfaces of devices. Therefore, the usual interfaces (remote controls, mice or keyboards) must be complemented by more accessible and natural interfaces such as a system of Automatic Speech Recognition (ASR) [1].

In this context, the CIRDO project[1] wherein the authors take part aims to promote autonomy and support of elderly people by caregivers through a social inclusion product. The objective of the project is to integrate an ASR system into this product to perform detection of distress situations, distress calls and voice commands. Such kind of voice based interaction is an emerging feature of many AAL related research projects [2, 3, 4, 5, 6] but this remains a very challenging area due to the atypical nature of the application (distant speech, aged people, noise, uncontrolled area, multi-speaker, etc.) [7].

One of the main challenges in this domain is to make sure that the ASR performance will be good enough to deliver a high quality voice order recognition system. This is a fear of the elderly population who are inclined to switch the system off if it has difficulties in understanding them. Most of the deployed ASR systems have reached a very good recognition rate in close, noise free talking, but their performances were rarely assessed with aged or children voice. A few studies compared ageing voice vs. non-ageing voice on ASR performance [8, 9, 10, 11], but their fields was quite far from our topic of home automation commands recognition. Moreover, an issue for our work was the non-existence of a speech corpus in French containing distress signals and automation commands.

The purpose of this study was to determine the impact of ageing voice on the ASR system performance and to find out which people characteristics might serve to predict ASR performance. The method we used is detailed in Section 3 after having discussed the related work in Section 2. Then the results of the evaluation are presented in Section 4 and an outlook on further work is given Section 5.

2. RELATED WORK

The perception of voice alteration with age has been the subject of many studies [12, 13, 14, 15, 16]. Elderly speakers are characterized by tremors, hesitations, imprecise production of consonants, broken voice, and slower articulation [13]. Regarding women, the changes seem partly due to an increase of the vocal cords mass due to some changing levels of certain hormones [17]. Regarding men, perception of gasp come from an incomplete closure of the vocal cords that would be compensated by an increasing tension in larynx [18]. From the 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 [19, 20].

Some studies have shown a significant increase in the standard deviation measures of the fundamental frequency of elderly, both men and women [21, 16, 15]. Stability of the fundamental frequency (F0) is reduced in elderly voice [12] and is associated with variability in the peak-to-peak amplitude of speech signal. Hesitations and gasping in pathological voices have been associated with increased noise in the speech signal driven by an aperiodic vibration of the vocal cords [22, 23]. Some measures of the ratio between noise energy and harmonics have quantified this phenomenon by comparing older and younger speakers [21, 24]. Incomplete closure of vocal cords was observed during vocalisation [14]. The study cited above [12] confirms fundamental frequency instabilities and the increasing noise on both sexes for healthy people with an average

[1] http://liris.cnrs.fr/cirdo/

SLPAT 2013, 4th Workshop on Speech and Language Processing for Assistive Technologies, pages 9–15,
Grenoble, France, 21–22 August, 2013.

age of 70. These studies show that aged voice presents a much greater variability than typical voice. Ability of state-of-the-art ASR systems to handle this kind of population can thus be questioned.

A more general study of Gorham-Rowan and Laures-Gore [12] highlights the effects of ageing on the speech utterance and the consequences on the speech recognition. The experiments carried out in automatic speech recognition have shown performance degradation for "atypical" population such as children or elderly people [25, 10, 26] and have shown the interests of an adaptation to the target populations [27, 26]. Speech recognition adapted to the voice of elderly people is still an underexplored area. The relevant languages are mainly English [10] and Asian languages such as Japanese [8]. A very interesting study [10] used some recordings of speeches delivered in the Supreme Court of the United States over a decade. These recordings are particularly interesting because they were used to study the evolution of recognition performance on the same person depending on his age over 7-8 years. These studies show that the performance of recognition systems decreases steadily with age, and that a special adaptation to each speaker can get closer to the scores obtained from the youngest speakers without adaptation. The implicit consequence is that the recognition system is adapted to a single speaker. To make the system adapted to the person, Renouard et al. [28] proposed to use the recognized words to adapt online the recognition models. Proposed in the context of home assistance, this research does not appear to have been pursued.

From an applicative point of view in the smart home context, speech recognition has been mainly implemented marginally in the field of voice commands in English. Indeed, most current studies use conventional sensors (presence sensor, door contact, etc.) and tactile interfaces (remote controls, handset, touch-screens) more reliable but less natural, offering fewer opportunities for interaction and comfort (for example: need to walk to reach the remote device). Among the advances in the field of voice controlled devices, a study conducted by Anderson [27] showed that a voice interface which is adapted (models acquired on 300 elderly speakers) allows to make voice requests on computer with the same performance than a query typed at the keyboard. This study has also revealed that only 2 of the 37 participants preferred the keyboard compared to the voice interface. In the same field, Kumiko [29] proposed a computer voice command interface that takes into account the possible sources of error (duration, intensity, vocabulary) to improve performance and feedback. While Interactive Voice Response is a pervasive component of today's telephone communication, some of which take into account the different voice population [30], voice control in smart home is clearly in its infancy. A large number of issues, such as noisy environment, number of sound sources (for example: several people), vocabulary coverage, coverage of speakers, etc. still need to be addressed [7]. Recently, Moir and Filho [31] proposed a low-coverage system using adaptive filters for a good recognition of keywords. But this research remains still exploratory.

To the best of our knowledge, no application of voice control in smart home has explicitly considered the problem of voice recognition of French elderly speakers, even though major advances in terms of ergonomics, safety and data acquisition with high semantic value can be made by this modality. From this short literature review, it can be emphasized that no study had considered French aged voice in smart home condition. Moreover, most studies considered the chronological age as global explanatory factor while many other effects can also be responsible for ASR performance degradation as raised by [11]. There is thus no certainty that age can predict the reliability of a voice-based control system. That is why our study includes an evaluation from the dependence perspective.

3. METHOD

To assess the impact of the ageing voice on ASR performance, we started by acquiring a corpus targeted to the elderly population. From this corpus and a non-aged one, the first task was to identify the most problematic phonemes and to check whether standard adaptation can be employed to reduce the discrepancy between aged and non-aged speakers at phoneme level. Once adapted, the second task was to assess whether measures other than strictly chronological age can explain ASR performance degradation.

3.1. Corpus collection

The corpus collection was performed sporadically from 2009 to 2012 in collaboration with a rehabilitation centre, volunteers and a nursing home. Targeted speakers were persons aged of more than 60 years old, able to read and with no mental disorder or pathologies altering the voice. The recording was done with a single microphone positioned about 30 cm from the speaker's mouth. Most speakers were sat, but some were in a wheelchair or laying in a bed. The recording was done using a computer and a home made software to prompt sentences to be read by the speaker and to record the utterances using voice activity detection. Given the targeted application (in-home voice commands and distress calls) the participants were requested to read a list of short distress/home automation and casual sentences such as *Aidez-moi* (Help me) or *Il fait beau* (It's sunny). Based on [32], who interviewed elderly people in nursing homes to identify and describe what situations of distress they could have experienced, we created a list of home automation orders the person could utter during a distress situation to request for assistance. Ten samples of each kind are given in Table 1.

The non-aged corpus was previously recorded in our laboratory in 2004 and was complemented in 2013 with sentences based on [32]. The procedure was similar to the aged corpus acquisition.

This aged and non-aged corpus is called the AD corpus (Anodin-Détresse: *anodin* means colloquial and *détresse* means distress).

Finally, another aged corpus, the ERES38 corpus (Entretiens RESidences 38: *Entretiens* means interviews) was acquired for model adaptation purpose. This corpus was recorded in 2011 in the living place of the person. During the interviews, we requested each speaker to read a text but they were also asked to talk freely about their life. The text was an article about gardening created by the experimenters in order to target phoneme issues reported in [9, 33].

All the corpora were annotated at the sentence level using the Transcriber software.

3.2. ASR system

The ASR toolkit chosen in our study was Sphinx3 [34]. This decoder used a context-dependent acoustic model with 3-state left-to-right HMM. The acoustic vectors are composed of 13 MFCC coefficients, the delta and the delta delta of each coefficient. This HMM-based context-dependent acoustic model was trained on the BREF120 corpus [35] which is composed of about 100 hours of annotated speech from 120 non-elderly

Sample	Distress Sentence	Home Automation Order	Casual Sentence
1	Aidez-moi !	e-lio appelle le samu !	Bonjour madame !
2	Au secours !	e-lio appelle les pompiers !	Ça va très bien.
4	Je me sens mal !	e-lio appelle les secours !	Ce livre est intéressant.
5	Je suis tombé !	e-lio appelle un docteur !	Il fait soleil.
3	Du secours s'il vous plaît !	e-lio appelle une ambulance !	J'ai ouvert la porte.
6	Je ne peux plus bouger !	e-lio appelle une infirmière !	Je dois prendre mon médicament !
7	Je ne suis pas bien !	e-lio appelle ma fille !	J'allume la lumière !
8	Je suis blessé !	e-lio appelle mon fils !	Je me suis endormi tout de suite !
9	Je ne peux pas me relever !	e-lio tu peux téléphoner au samu ?	Le café est brûlant !
10	Ma jambe ne me porte plus !	e-lio il faut appeler les secours !	Où sont mes lunettes ?

Table 1: Examples of sentences of the AD corpus

French speakers. We called it the generic acoustic model.

3.3. Language model

A general language model (LM) was estimated from the French *Gigaword* corpus which is a archive of newswire text data that has been acquired over several years by the Linguistic Data Consortium (LDC) at the University of Pennsylvania[2]. It was 1-gram with 11018 words. Moreover, to reduce the linguistic variability, a 3-gram domain language model was learned from the sentences used during the corpus collection described in Section 3.1, with 88 1-gram, 193 2-gram and 223 3-gram models. Finally, the language model was a 3-gram-type which results from the combination of the general language model (with a 10% weight) and the domain one (with 90% weight). This combination has been shown as leading to the best WER for domain specific application [36]. The interest of such combination is to bias the recognition towards the domain LM but when the speaker deviates from the domain, the general LM makes it possible to correctly recognise the utterances.

3.4. Word error rate and phoneme matching

The simplest and most common way to evaluate ASR performances is to compute the Word Error Rate (WER). The WER is computed by first aligning the output (the decoded speech) with the reference (i.e., the ground truth) and then applying $WER = \frac{I+D+S}{N}$ where I, D and S is the number of insertions, deletion and substitution of words and N is the number of words in the reference.

Though this measure was used in many related studies [8, 10, 11], it does not indicate which specific phonemes play a role in the ASR performance degradation. To do so, the annotation should be performed at the phoneme level. However, this is a very laborious and time-consuming task which furthermore requires a good level of expert agreement. That is why we analysed the results of the forced alignments. The forced alignment algorithm that was used is the one of Sphinx3.

Forced alignment consists in finding the boundaries of phonemes in an utterance knowing the uttered sentence. This sentence is mapped in phoneme (using a dictionary) which is used to constrain an optimal alignment between the acoustic model and the speech utterance. The forced alignment scores are for each signal segment within a boundary, the likelihood of belonging to a phoneme model. This score can be interpreted as a proximity to the "standard" pronunciation, modelled by the generic acoustic model. The differences in scores of phoneme categories between the aged group and the non-aged group allow to identify which phonemes are the most difficult for the ASR system. We are not aware of any study having used such method to assess ASR performances.

[2]http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T17

Unvoiced Plosive	p, t, k
Voiced Plosive	b, d, g
Nasal Consonant	m, n, ŋ, ɲ
Liquid Consonant	l
Unvoiced Fricative	f, s, ʃ
Voiced Fricative	v, z, ʒ, ʁ
Front Vowel	i, e, ɛ
Central Vowel	y, ø, œ, ə
Back Vowel	u, o, ɔ
Open Vowel	a, ɑ
Nasal Vowel	ɛ̃, ɑ̃, ɔ̃, œ̃
Semi-Vowel	ɥ, j, w

Table 2: Phoneme categories (IPA symbols)

Phonemes were grouped according to their highest level catgories as shown in table 2.

3.5. Adaptation with MLLR

Once the phonemes are identified, the most common method to overcome the ASR limitation is to apply speaker adaptation. Speaker adaptation consists in generating a new acoustic model from a generic one and some new annotated speech in limited quantity. One of the most popular technique is to apply the Maximum Likelihood Linear Regression (MLLR) which is particularly adapted when a limited amount of data per class is available. MLLR is an adaptation technique that uses small amounts of data to train a linear transform which warps the Gaussian means so as to maximize the likelihood of the data. The principle is that acoustically close classes are grouped and transformed together.

3.6. Assessing the level of autonomy

Despite the acoustic adaptation, there might be a disparity between the WERs of the elderly group even in aged people of the same age category. Therefore, we investigated other criteria and focused on elderly dependence. As reference, we used a French national test which is daily used in assessing the degree of loss of autonomy: the AGGIR (Autonomie Gérontolo-

gie Groupes Iso-Ressources) grid[3]. The degree of autonomy loss is evaluated in terms of physical and cognitive decline. According to the result of this test, the person can receive financial support: the Personalized Allocation of Autonomy (APA). The evaluation is done using 17 variables. Ten variables refer to the loss of physical and cognitive autonomy: coherence, orientation, washing, feeding, disposal, transfers (to rise, to lie down, to sit down), internal displacement, external displacement and remote communication. Seven variables refer to the loss of domestic and social autonomy: personal management of budget and possessions, cooking, cleaning, transporting, purchasing, treatment monitoring and past time activities. Each variable is coded with A (independent), B (partially dependent) and C (totally dependent). The GIR (Iso-Ressources Group) score is computed from the variables to classify the person in one of the six groups: GIR 1 (total dependence) to GIR 6 (total autonomy).

4. RESULTS

4.1. Collected Corpus

4.1.1. The AD80 French test corpus

The AD corpus (cf. 3.1) was acquired from 95 speakers (36 men and 59 women) which are divided into two groups: the elderly group composed of 43 speakers (11 men and 32 women), 62 to 94 years old, with 2796 distress and home automation sentences for a duration of 1 hour 5 minutes, and 3006 casual sentences for a duration of 1 hour 6 minutes, and the non-elderly group composed of 52 speakers, 18 to 64 years old, with 3903 distress and home automation sentences for a duration of 1 hour 18 minutes, and 3897 casual sentences for a duration of 1 hour 12 minutes.

We fixed the limit of the non-aged group at 65 years old, but we recorded 2 people aged 62 and 63 years old with autonomy loss, looking very aged physically and living in nursing home. Thus we included this two persons, as exceptions, in the aged group.

For the 43 speakers of the aged AD corpus, a GIR score was obtained after clinicians filled the AGGIR grid.

Finally, the AD corpus is made up of 13,602 annotated sentences, with 4 hours and 42 minutes of recording.

4.1.2. Collection of the training ERES38 corpus

The ERES38 (cf. 3.1) corpus was acquired from 22 elderly people (14 women and 8 men) between 68 and 98 years old. The corpus included 48 minutes of read speeches (around 2 minutes per speaker) and 17 hours of interviews. The speakers lived in specialized institutes, such as nursing homes and were cognitively intact without severe disabilities.

4.2. Phoneme distance between aged and non-aged voice

When performing ASR using the generic acoustic model on the distress/home automation sentences of the AD corpus, we obtained an average WER of 9.07% for the non-elderly group, and an average WER of 43.47% for the elderly group. Thus, we observed a significant performance degradation of ASR for elderly speech, with an absolute difference of 34.40%. Figure 1 represents the WER according to the chronological age for both groups. It shows that the WER is globally higher for elderly group as previous studies showed [8, 10, 11]. However, it can also be seen that the variability between speakers also increases with the age. For instance, some 83 years old speakers have their WER ranging from 13.6% to 80.2%. Standard deviation is 6% for the non-elderly group and 17.27% for the elderly group. In other words, the WER is far less predictable in the elderly group than in the non-elderly group. Consequently, we have to deal with the fact that a speech recognition with such a system can work very well with some of the elderly speakers, and very badly with others.

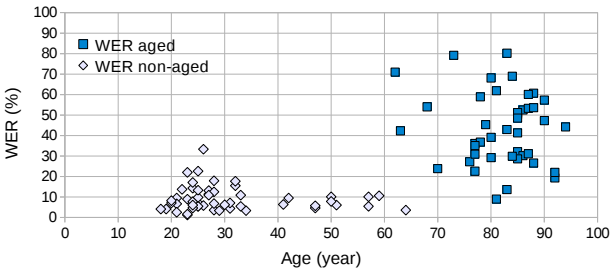

Figure 1: WER as a function of age for aged and non-aged groups

The forced alignment scores on both AD groups non-elderly and elderly with the generic acoustic model are presented in Figure 2 based on phonemic categories.

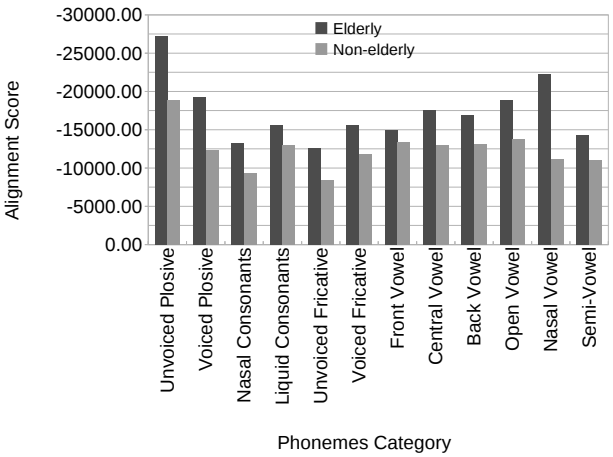

Figure 2: Forced alignment scores by phonemes categories before adaptation

The relative differences of forced alignment scores observed between both groups led to sort the phoneme categories in descending order of differences: nasal vowels (-100,34%), voiced plosives (-56,55%), unvoiced fricatives (-50,48%), unvoiced plosives (-44,05%), nasal consonants (-41,03%), open vowels (-37,12%), central vowels (-34,80%), voiced fricatives (-31,30%), back vowels (-29,26%), semi-vowels (-29,18%), liquids (-19,99%), and front vowels (-11,89%). The repartition of French phonemes inside the different groups are presented in Table 2.

For the elderly group, the alignment scores are lower than those obtained for the non-elderly group especially for plosives and nasal vowels. Based on the relative differences, the phoneme categories most affected for elderly group are nasal vowels, plosive consonants, unvoiced fricatives and nasal consonants.

[3]http://vosdroits.service-public.fr/F1229.xhtml

4.3. Impact of the acoustic adaptation on ASR performance with aged voice

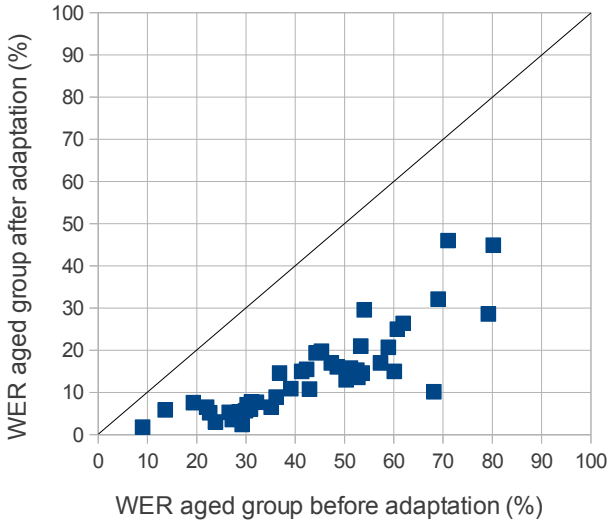

Figure 3: WER on aged group before and after adaptation

Figure 3 shows that using the MLLR adapted acoustic model was able to reduce the WER significantly for all speakers of the AD corpus. With the global MLLR adaptation using ERES38, the average WER was 14.52%. Compared to the 43.47% WER without adaptation (see Section 4.2), the absolute difference was -28.95%. Furthermore, the speaker with the worst performance had his error rate reduced from 80.2% to 44.9%, and the speaker with the best performance had his error rate reduced from 9% to 1.8%. Also, the standard deviation was reduced from 17.27% to 10.34%, showing a reduction of the variability between the speakers.

A comparison between the forced alignment scores obtained for non-elderly without adaptation and for elderly after adaptation using the ERES38 corpus is shown in Figure 4. On the whole, the scores for the elderly after adaptation are better than those of non-elderly with the generic acoustical model.

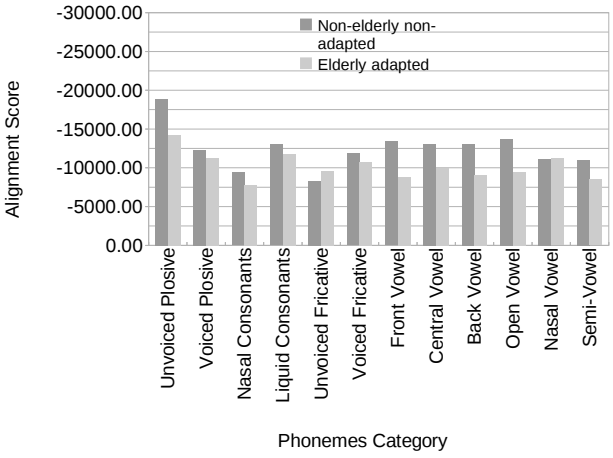

Figure 4: Forced alignment scores by phonemes categories for non-elderly with the generic acoustical model and for elderly after adaptation

Indeed, the use of an acoustical model adapted to elderly people reduces the mismatching of phonemes. The alignment scores of the adapted model presented in Figure 4 show that the average distance has reduced below the non-aged one for all phonemes except for unvoiced fricatives and nasal vowels.

From an applicative point of view, this test shows that we can use a database of elderly speech in MLLR adaptation with speakers which are different from the test database. Even though the size of the corpus is small, we have a significant improvement of WER. Furthermore, this demonstrates that the voices of ageing people have common characteristics.

4.4. Influence of elderly dependence on ASR system

Despite the acoustic adaptation, there is a great variability between the WERs of the elderly group. Therefore, we investigated to establish if the level of elderly dependence can be an indicator of the ASR performance for the elderly group. Figure 5 shows a box-and-whisker diagram of the WER from MLLR adaptation as a function of the elderly dependence. Four speakers were in GIR 2, two speakers were in GIR 3, 21 speakers were in GIR 4, one speaker was in GIR 5 and 15 speakers were in GIR 6. No speaker was represented in GIR 1. Due to the small number of speakers in GIR 2, GIR 3 and GIR 5, we merged GIR 2 with GIR 3 and GIR 4 with GIR 5.

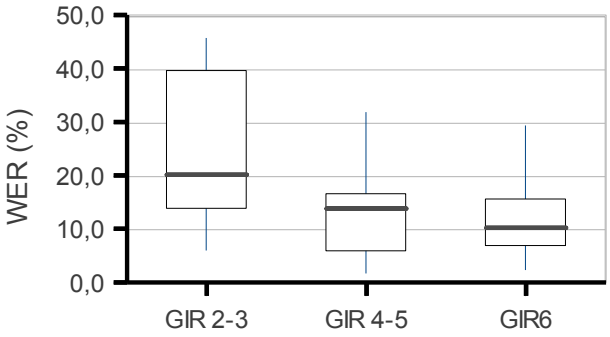

Figure 5: WER as a function of levels of dependence

From Figure 5 it can be seen that WERs are different according to the GIR category. Indeed, the WER averages for GIR 2-3, GIR 4-5 and GIR 6 are respectively 25.2%, 13.2% and 12.2%, and the WER standard deviations are respectively 16.8%, 8.4% and 7.6%. Then, we performed an ANOVA test on the groups GIR 2-3, GIR 4-5 and GIR 6. From this test, the GIR score have a significant effect on WER ($F(2, 40) = 4.3; p < 0.05\%$).

We conducted a Bonferroni post-hoc analysis to characterize which groups were significantly different from other groups. The post-hoc test highlighted that there was a significant difference between the GIR 2-3 group and both groups GIR 4-5 and GIR 6, while there is no significant difference between GIR 4-5 and GIR-6.

5. CONCLUSION

The paper presents our study on the behavior of an ASR system with elderly voices. Given the absence of a corpus containing the voice of elderly in French language usable for testing ASR system, we recorded the AD corpus. From this corpus, we observed an increase of the average WER of the ASR system for elderly people, with an absolute difference between non-elderly

and elderly voice of 34.4%. With forced alignment, we analyzed which phonemes for elderly speech were posing the most problems to ASR systems. These results allowed us to proceed to the recording of the ERES38 corpus, allowing us to adapt the generic acoustic model to the voice of elderly people through the MLLR adaptation method. The global MLLR adaptation was interesting because with less than one hour of recordings from speakers different from the test speakers, we obtained a WER close to the case of recognition with the generic acoustic model on non-elderly group, with a WER of 14.53%, against 43.47% before adaptation. Moreover, we showed that inside the elderly group, the WER was not correlated with the age but could be correlated with the level of dependence due to a general physical degradation. The continuation of our work would be to show how the different parameters of the AGGIR grid are correlated to the WER. Therefore, predicting the ASR behavior would allow in facilitating the use of these new technologies in the daily life of the dependent elderly people.

6. ACKNOWLEDGMENT

This study was funded by the National Agency for Research under the project CIRDO - Industrial Research (ANR-2010-TECS-012). The authors would like to thank to Mrs Vézignol and Bonnefond-Jimenez, Mr Debrus, Mrs Aman, Bron, Lalande and Martins of the medical institutions SSR "Les Cadières" and EHPAD "Château de Labahou" for their help in corpus recording. Special thanks to R. Dugheanu, J. Le Grand and Y. Sasa for their active contribution, and to various elderly and caregivers who agreed to participate in the recordings.

7. References

[1] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly — Acceptability and objection aspects," *Personal and Ubiquitous Computing*, vol. 17, no. 1, pp. 127–144, 2013.

[2] J.-C. Wang, H.-P. Lee, J.-F. Wang, and C.-B. Lin, "Robust environmental sound recognition for home automation," *IEEE Transactions on Automation Science and Engineering*, vol. 5(1), p. 25–31, 2008.

[3] A. G. Jianmin Jiang and S. Zhang, "Hermes: a FP7 funded project towards computer aided memory management via intelligent computations," in *3rd Symposium of Ubiquitous Computing and Ambient Intelligence*, 2009, p. 249–253.

[4] O. Brdiczka, M. Langet, J. Maisonnasse, and J. Crowley, "Detecting human behaviour models from multimodal observation in a smart home," *IEEE Transactions on Automation Science and Engineering*, vol. 6(4), p. 588 –597, 2009.

[5] P. Milhorat, D. Istrate, J. Boudy, and G. Chollet, "Hands-free speech-sound interactions at home," in *20th European Signal Processing Conference (EUSIPCO)*, 2012, pp. 1678–1682.

[6] M. Vacher, F. Portet, B. Lecouteux, and C. Golanski, *Telhealthcare Computing and Engineering: Principles and Design*. CRC Press, Taylor and Francis Group, 2013, no. 21, ch. Speech Analysis for Ambient Assisted Living: Technical and User Design of a Vocal Order System, pp. 607–638, ISBN: ISBN-978-1-57808-802-7.

[7] M. Vacher, F. Portet, A. Fleury, and N. Noury, "Development of Audio Sensing Technology for Ambient Assisted Living: Applications and Challenges," *International Journal of E-Health and Medical Communications*, vol. 2, no. 1, pp. 35–54, 2011.

[8] A. Baba, S. Yoshizawa, M. Yamada, A. Lee, and K. Shikano, "Acoustic models of the elderly for large-vocabulary continuous speech recognition," *Electronics and Communications in Japan, Part 2*, vol. 87, pp. 49–57, 2004.

[9] R. Privat, N. Vigouroux, and P. Truillet, "Etude de l'effet du vieillissement sur les productions langagières et sur les performances en reconnaissance automatique de la parole," *Revue Parole*, vol. 31-32, pp. 281–318, 2004.

[10] R. Vipperla, S. Renals, and J. Frankel, "Longitudinal study of ASR performance on ageing voices," *Interspeech*, pp. 2550–2553, 2008.

[11] T. Pellegrini, I. Trancoso, A. Hämäläinen, A. Calado, M. S. Dias, and D. Braga, "Impact of Age in ASR for the Elderly: Preliminary Experiments in European Portuguese," in *Advances in Speech and Language Technologies for Iberian Languages - IberSPEECH 2012 Conference, Madrid, Spain, November 21-23, 2012. Proceedings*, 2012, pp. 139–147.

[12] M. Gorham-Rowan and J. Laures-Gore, "Acoustic-perceptual correlates of voice quality in elderly men and women," *Journal of Communication Disorders*, vol. 39, pp. 171–184, 2006.

[13] B. Benjamin, "Frequency variability in the aged voice," *Journal of Gerontechnology*, vol. 36, pp. 722–726, 1981.

[14] S. Linville and E. Korabic, "Elderly listeners' estimates of vocal age in adult females," *Journal of the Acoustical Society of America*, vol. 80, pp. 692–694, 1986.

[15] E. Morgan and M. Rastatter, "Variability of voice fundamental frequency in elderly female speakers," *Perceptual and Motor Skills*, vol. 63, pp. 215–218, 1986.

[16] R. Morris and W. Brown, "Age-related differences in speech variability among women," *Journal of Communication Disorders*, vol. 27, pp. 49–64, 1994.

[17] I. Honjo and N. Isshiki, "Laryngoscopic and voice characteristics of aged persons," *Archives of Otolaryngology*, vol. 106, pp. 149–150, 1980.

[18] W. Ryan and K. Burk, "Perceptual and acoustic correlates in the speech of males," *Jounal of Communication Disorders*, vol. 7, pp. 181–192, 1974.

[19] N. Takeda, G. Thomas, and C. Ludlow, "Aging effects on motor units in the human thyroarytenoid muscle," *Laryngoscope*, vol. 110, pp. 1018–1025, 2000.

[20] P. Mueller, R. Sweeney, and L. Baribeau, "Acoustic and morphologic study of the senescent voice," *Ear, Nose, and Throat Journal*, vol. 63, pp. 71–75, 1984.

[21] S. Xue and R. Deliyski, "Effect on aging on selected acoustic voice parameters: Preliminary normative data and educational implications," *Educational Gerontology*, vol. 27, pp. 159–168, 2001.

[22] L. Eskenazi, D. Childers, and D. Hicks, "Acoustic correlates of vocal quality," *Journal of Speech and Hearing Research*, vol. 33, pp. 298–306, 1990.

[23] J. Selby, H. Gilbert, and J. Lerman, "Perceptual and acoustic evaluation of individuals with laryngopharyngeal reflux pre- and post-treatment," *Journal of Voice*, vol. 17, pp. 557–570, 2003.

[24] C. Ferrand, "Harmonic-to-noise ratio: An index of vocal aging," *Journal of Voice*, vol. 16, pp. 480–487, 2002.

[25] J. Wilpon and C. Jacobsen, "A study of speech recognition for children and the elderly," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 1996, pp. 349–352.

[26] M. Gerosa, Giuliani, and F. D., Brugnara, "Towards age-independent acoustic modeling," *Speech Communication*, vol. 51(6), pp. 499–509, 2009.

[27] S. Anderson, N. Liberman, E. Bernstein, S. Foster, E. Cate, B. Levin, and R. Hudson, "Recognition of elderly speech and voice-driven document retrieval," in *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '99*, vol. 1, 1999, pp. 145–148.

[28] S. Renouard, M. Charbit, and G. Chollet, *Independent Living for Persons with Disabilities*, 2003, ch. Vocal interface with a speech memory for dependent people, pp. 15–21.

[29] O. Kumiko, M. Mitsuhiro, E. Atsushi, S. Shohei, and T. Reio, "Input support for elderly people using speech recognition," IEIC Technical Report (Institute of Electronics, Information and Communication Engineers), Tech. Rep. 104(139), 2004.

[30] E. Pinto, D. Charlet, H. François, D. Mostefa, O. Boëffard, D. Fohr, O. Mella, F. Bimbot, K. Choukri, Y. Philip, and C. F., "Development of new telephone speech databases for french : the neologos project," in *4th International Conference on Language Resources and Evaluation*, 2004, pp. 1–4.

[31] T. Moir and G. Filho, "From science fiction to science fact: A smart-house interface using speech technology and a photo-realistic avatar," in *15th International Conference on Mechatronics and Machine Vision in Practice*, 2008, pp. 327–333.

[32] M.-E. B. Chaumon, B. Cuvillier, S. Bouakaz, and M. Vacher, "Démarche de développement de technologies ambiantes pour le maintien à domicile des personnes dépendantes : vers une triangulation des méthodes et des approches," in *Actes du 1er Congrès Européen de Stimulation Cognitive*, Dijon, France, 23-25 May 2012, pp. 121–122.

[33] F. Aman, M. Vacher, S. Rossato, R. Dugheanu, F. Portet, J. le Grand, and Y. Sasa, "Etude de la performance des modèles acoustiques pour des voix de personnes âgées en vue de l'adaptation des systèmes de RAP (Assessment of the acoustic models performance in the ageing voice case for ASR system adaptation) [in French]," in *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 1: JEP*, 2012, pp. 707–714.

[34] K. Seymore, C. Stanley, S. Doh, M. Eskenazi, E. Gouvea, B. Raj, M. Ravishankar, R. Rosenfeld, M. Siegler, R. Stern, and E. Thayer, "The 1997 CMU Sphinx-3 English broadcast news transcription system," in *DARPA Broadcast News Transcription and Understanding Workshop*, Lansdowne, VA, USA, 1998.

[35] L. Lamel, J. Gauvain, and M. EskEnazi, "BREF, a large vocabulary spoken corpus for french," in *Proceedings of EUROSPEECH 91*, vol. 2, Geneva, Switzerland, 1991, pp. 505–508.

[36] B. Lecouteux, M. Vacher, and F. Portet, "Distant Speech Recognition in a Smart Home: Comparison of Several Multisource ASRs in Realistic Conditions," in *Interspeech 2011*, Florence, Italy, 2011, p. 4p.