Elderly Speech Collection for Speech Recognition Based on Crowd Sourcing

Ana Júdice[1], João Freitas[1], Daniela Braga[1,2], António Calado[1], Miguel Dias[1,2], António Teixeira[3], Catarina Oliveira[4]

[1]Microsoft Language Development Center, Portugal

[2]ADETTI – ISCTE, IUL, Lisbon, Portugal

[3]Dep. Electrónica, Telec & Informática (DETI)/IEETA - Universidade de Aveiro,

[4]Catarina Oliveira, Escola Superior de Saúde - Universidade de Aveiro,

{a-anju, i-joaof, i-dbraga, i-antonc, Miguel.Dias}@microsoft.com, ajst@ua.pt, coliveira@ua.pt

Abstract: - In this paper, a preliminary experiment on elderly speech data collection based on a crowd-sourcing approach is described. From this experiment, a usability study of the web platform used for this collection was conducted and a few changes were made in order to meet senior users' needs. The main goal of this work is to collect as much speech as possible from senior citizens in order to improve the current speech recognition accuracy rates, which does not perform well with speech uttered by the elderly. This work is under the scope of a larger project[1] focusing on the development of technologies and services to support healthy, productive and active citizens, using multimodal interfaces and particularly embedded speech technology.

I. INTRODUCTION

Technology applications are often difficult to use for elderly people, either because they are not adapted to elderly's specific needs, or because they have unfriendly interfaces in most cases, which contribute to their social isolation and exclusion. In terms of usability, it has been demonstrated that speech is the easiest and most natural way for human-machine interaction [1] and that this type of interface is often preferred, especially when mobility is reduced or the screen is too small for a touch interface.

Although speech interfaces are becoming more popular and covering more languages, the ASR (Automatic Speech Recognition) systems are still not adapted to deal with the voice of all population segments, specifically the ageing population, because: a) ASR acoustic models are often not trained with enough elderly speech, and b) senior citizens have not been seen as a relevant market sector for ICT related applications.

However, the growing senior population in Europe is worrying the European Commission, who is putting a great effort in funding Assistive Technology and Ambient Assisted Living solutions for ageing well and independent living. In a scenario of fast population ageing like the one we are facing, European policies are encouraging seniors to go online and to use technology in order to improve their quality of life and reduce their dependency on caregivers and social institutions. In this scenario, speech and multimodal interfaces must be available and usable by all, especially by the senior population.

Our voice characteristics change as we get older and that has a significant impact in speech technology performance. Current Speech Recognition systems are trained with hundreds of hours of adults' speech, but they often don't include elderly speech, which means the acoustic models are not prepared to recognize elderly speech voice. Recognizing elderly speech is even more challenging in certain languages (like Portuguese), due to the lack of speech training data. In section II, a discussion of what is elderly speech and a characterization of the senior population in Portugal is presented. In section II, it is also debated the lack of research and development on speech technology for elderly users in contrast with the growing need for this type of technology, especially embedded in Assistive Technology solutions (also described in section II) when we consider the fast growth of senior population.

It is well known that ASR systems based on statistical methods require vast amounts of transcribed speech data in order to achieve acceptable accuracy rates. Acquiring a lot of speech data is particularly difficult when addressing less-resourced languages or even any other language that is not amongst the big five (English, Spanish, French, German and Italian) in terms of market economic relevance. The main reason for this is that these corpora are expensive and recruiting speakers has proven to be quite costly and hard to manage. Besides, some speech databases lack quality because of the bad recording conditions, sample rates inconsistency, erroneous, inconsistent or inexistent transcription, etc.[2]. This paper describes

[1] This work was done on the scope of the Living Usability Lab project (http://www.livinglab.pt), under the QREN program, co-funded by FEDER through COMPETE program. The authors would also like to thank to the senior speakers who kindly donated their voice.

an ongoing work to tackle this issue by using a crowd-sourcing approach (explained in Section III).

Besides speech technology, design of web interfaces for elderly people was another motivation for this work. When preparing a platform to collect elderly speech to improve speech recognition rates, we found that the web interface we had used in previous projects had to be adapted to this population segment. In sections IV and V, that experience is reported and discussed. In section VI, we describe the changes we made to the initial web platform in order to adapt it to the seniors' users.

II. State of the Art and Background Information

A. What is Elderly Speech?

Until now, many disciplines, such as Psychology, Social Psychology, Sociology and Gerontology, have dealt with the subject of elderly people in various ways. The topic of elderly speech has established itself as an interdisciplinary research area with diverse approaches of investigation, which can be proved by the disparity of studies and methodologies, which are, in most of the cases, extremely difficult to compare.

Literature draws a divergent picture about how to characterize elderly speech. While some propose criteria to distinguish between elderly versus teenagers' or adults' speech, there are others denying that there are clear-cut differences [3]. The absence of a single deterministic phonetic cue, existent, for example, in gender determination, makes elderly speech classification inexact. Since aging increases the difference between biological age and chronological age and considering that biological aging can be influenced by factors such as, abuse or overuse of the vocal folds, smoking, alcohol consumption, psychological stress/tension, or frequent loud/shouted speech production without vocal training [4][5] it is not possible to determine an exact age limit for speech to be considered as elderly.

Conducted studies consider ages between 60 and 70 as the minimum age for the elderly age group [6]. Fiehler [7] follows the option that what is hastily called to be 'typical for the speech of elderly people' results from different situational circumstances, from which different registers are drawn, being this with respect to lexical and grammatical aspects. Though in contrast to teenagers' speech, which is often used as a social identifying characteristic, seniors do not look for acceptation of a group in the same way, because of their experience of life, their (acquired) social status, and at the same time, because they are not necessarily dependent of a linguistic assignment from a peer-group.

Observations considering the voice of elderly people have proved that it is possible to state differences between elderly speech and teenagers or adults speech on an acoustic phonetic level [8]. With increasing age there is a deprivation of chest voice, general changes in frequencies, in the voice quality and the timbres. Changes in the heights of vowel formant frequencies particularly appear in older men, not only for biological reasons, but also because of social changes.

According with [8], [9], the speech rate is also slower. Simultaneously more breaks, more speech errors and a lower volume of speech were detectable. American studies [10] also conclude that elderly subjects in general produce fewer morphemes per utterance as well as fewer utterances per minute. Additionally, they assume that the subjects, while aging, eliminate more often compulsory grammatical morphemes as well as articles and possessive pronouns. Many studies agree that utterances get overall shorter with increased age, that seniors produce less correct verb tenses and also other correct morphological forms. It was also reported in [10] an age-dependent reduction of complex syntactic structures concerning written language.

B. Automatic Speech Recognition of Elderly Speech

Age is a key physiological characteristic of a speaker that must be considered in human-computer interfaces (HCIs) based on speech [12]. Although being a stable characteristic when compared with the awareness and emotional state of a speaker, age influences the performance of a SR engine, as several parameters of the speech wave form are modified, such as fundamental frequency, jitter, shimmer and harmonic noise ratio [13]. Additionally, with age, the cognitive and perceptual abilities decrease [14][15]. Studies show that SR accuracy of subjects aged below 15 or above 70 decreases dramatically, when using acoustic models specifically trained with young to middle aged adults. Experiments demonstrate a word error rate (WER) increase of 50% when comparing senior users with a middle age user group [16]. On the other hand, in [17], it was found that SR engines trained with elderly specific acoustic data showed a significant decrease (around 12% better) in WER when compared to engines trained with regular data. Similar results were obtained in [18] where an improvement of 2.9% in the WER was achieved when training an acoustic model with an elderly only corpus.

In [19], the WER increases 11% when the test corpus is entirely formed by speech from an older group, in relation to a test corpus formed exclusively by speech from the young group.

In brief, generic trained ASR systems perform significantly worse when used by the elderly

population, due to various factors. The typical strategy to improve ASR performance under these cases is to collect speech data from elderly users in the specific domain of the target application and train elderly-only acoustic models.

C. Speech Interfaces for Assistive Technologies and applications

The existent companies and organizations that develop applications for seniors usually use elderly targeted places such as, nursing homes, government subsidized housing, retirement communities, senior's centers and public libraries to develop, test and disseminate their software. The main objective of these applications is to provide access to web contents, to enable communication with friends and family, promote literacy, and overcome elderly reluctance towards media and electronic devices. According to [20], the number of seniors connected to the internet is rising hastily being the fastest-growing demographic group online. Internet is also increasingly becoming an important resource for information about health and health care options, communication and news. By being connected to the outside world, senior citizens become more socially integrated and have fewer depressive symptoms. The applications for seniors are characterized by friendly interfaces with buttons and font sizes above normal in order to tackle with low vision issues that are characteristic of the population in this age group. As stated in 0 speech can also be applied as an HCI in software that targets the elderly age group. Also, when used by elderly, speech interfaces are considered to be natural and practical to use [22].

When looking at the available speech-enabled solutions such as, Windows 7 accessibility features, QualiWorld platform [23], NIHSeniorHealth Website [24], Verbose Text to Speech [25], I2net Orion [21], etc., common features characteristics can be found. These applications focus on delivering to the user a simpler and easier access to daily tasks such as, using the internet. Amongst the common features are: Screen readers by using a TTS system; speech commanding and dictation to interact not only with the computer but also with household appliances; scaling the user interface to make text and graphics more readable; changing shapes, contrast and sizes of pictures; etc.

D. Characterization of the elderly population in Portugal

According to recent census and studies [26], the demographic evolution in Portugal is showing a continuous ageing, mainly because people live longer and the birth rate is declining. Since the beginning of the decade until 2008, the average life expectancy increased 2.26 years on average for both sexes – 2.46 years for men and 2.05 years for women. The estimated life expectancy in the period 2006-2008 is 75.49 years for men and 81.74 years for women. The overall birth rate used to be 1.56 children per woman in 2000, but that number declined to 1.33 children per woman in 2007, the lowest figure recorded thus far in Portugal. On December 2008, the elderly population in Portugal (65 and above) was 17.6% already. In 2004, the projections made in [27] showed that the increase of the population ageing is a phenomenon observed equally in all regions of the country. In general, an increase between 63.2% and 76.5% is to be expected. The same document predicts that the ageing rate will continue to increase until 2050, whichever scenario might be chosen. In the worst case scenario, the ageing index could reach to a point of 398 elderly (65 years and above) for every 100 youths in 2050, thus quadrupling the value of the ageing index in the course of 50 years. Even in the base scenario, the prediction is 243 elderly for every 100 youths, whereas the most optimistic scenario limits the number of elderly people per 100 youths to 190.

From the perspective of the WHO (World Health Organization), it is paramount that policies and programs to keep the older population active and productive are defined, so as to guarantee the sustainability of the current society. Maintaining an active aging is one of the biggest challenges our society will have to face in the coming decades in order to promote equal opportunities for all. In 2008 the working population rate continued to rise [28], which was mainly caused by the growing number of women joining the labor market, by the postponing of the retirement age and by the dynamics of migration flows. This increase of the working population rate also corresponded to a higher level of education and higher level of workforce qualification. Between 1998 and 2008, an increase in the total working population of around 529000 individuals corresponded to an increase of about 685000 individuals having at least completed secondary education, being noted that this relationship has intensified between 2004 and 2008 [29]. However, the proportion of assets with the corresponding level of education to higher education remained relatively low, standing at 14.8% in 2008.

As for as the usage of TIC is concerned, keeping the gender and the age group of the individuals in mind, we can conclude [30] that the computer and internet users are predominantly male, although there is a slight difference compared with use by female users. With regard to the age of users, note that the vast majority it is understood between 16 and 24 years, with one big difference from the older age (65-74), the latter being the least used computer and internet (8.1% and 6.6% respectively). We also can conclude that the

use of these technologies is directly linked to education level, and most users are located in higher education, followed by individuals with secondary education and, finally, individuals with lower education level (up to the 3rd cycle). The employment status is also a motivating factor to using these technologies, and most students who use computer (99.3%) and internet (96.7%), followed by the employee, unemployed, and finally, the inactive, with only 15% of computer usage and 12.1% of Internet usage.

III. The Crowd-Sourcing concept

Crowd-sourcing is a term used to describe the leveraging of vast amounts of people to achieve a specific goal in a collaborative manner over the Internet. Many crowd-sourcing initiatives have been made possible due to the availability of Web 2.0 technologies, which enable massive collaboration projects to take place. Crowd-sourcing can be considered as a distributed process for the resolution of problems. Typically the process is as follows: an entity has a problem and needs to solve it in a cost effective way. The entity publishes the problem in the web and usually provides the tools to solve it. Users (the crowd) respond to the call and propose solutions to the problem. The publishing entity chooses the winning solution and rewards the user/users accordingly. Rewards vary from money incentives to just public recognition. The publishing entity will own the final winning solution. Multiple solutions can be found across the web in order to digitize old books[31], transcribe speech[32], classify tunes[33], classify galaxies from the Sloan sky survey[34], find ideas for proposed problems[35], image[36] and video [37] tagging and even build a summary of the entire Human knowledge [38].

For the elderly speech data collection described in this paper, we used the Doaravoz, a different version of the Yourspeech platform [39], adapted to senior users. Doaravoz is a tool based on crowd sourcing approaches [40], designed to collect speech data at negligible costs for any language. The concept behind this system is to provide the user with an entertainment reward in exchange for his/her speech. In a previous collection done using the Yourseech platform, we invited Portuguese speakers to aid in the development of new ASR technology for European Portuguese. 25 hours of pure speech (with no silences or pauses) were collected showing that YourSpeech was a viable way for obtaining speech data at marginal costs given the fact that appropriate marketing and advertisement actions are taken. The preliminary collections reported in this paper show that a few aspects in the prior Yourspeech platform had to be changed, in order to adapt it to senior users, as described in section IV, V and VI.

IV. Elderly Speech Data Collection: preliminary tests

A. Sample group

The sample group selection was based on the following requirements: subjects should be over 65 years old, because that is the age group that is usually not present in the common speech data collections for Speech Recognition (such as SpeechDat [41]) and should be able to read, to hear, and to speak without any obvious voice pathology or disability. The sessions took place in Aveiro and in Lisbon, in April 2010. The first population sample consisted of 2 active female subjects, aged 75 and 76, from the Lisbon area, with a high level of education (secondary education and PhD). The second sample group was from Aveiro and was composed by 8 people over 65 years old (4 men and 4 women); 7 participants were at a public nursery house and one was a nursery house assistant. The 7 participants were over 80 years old, lived most of their lives in rural areas and had primary education level.

B. Description of the trial data collection session

In this preliminary data collection, all recording sessions were monitored by an assistant who guided the user and registered future improvements based on user experience and feedback. Then a preliminary test was conducted (to assess reading skills and hearing capabilities) in order to determine the user's eligibility for the complete trial. A brief questionnaire to collect speaker information was also conducted. Then the user was invited to record 50 prompts using the Yourspeech data collection platform (hereafter Doaravoz 1.0) installed on a local PC and using a headset. The sessions took place in a silent room. Each recording session lasted roughly 30 minutes, except for one subject, who took longer.

A screenshot of the Doaravoz 1.0 data collection platform in action can be seen in Figure 1:

Fig. 1. Screenshot of the Doaravoz 1.0.

The data collection platform consists of an HTTP server and client application that performs the actual audio recording using an ActiveX control via Internet explorer. The architecture of the platform operates on the internet (in http://www.doaravoz.com), but in the specific case of this trial, it ran entirely in stand-alone mode in a laptop. Audio files were recorded at 16 KHz, 16 bits, no compression and in wave format.

The used prompts were extracted from public European Portuguese corpora such as Cetem Público and Portuguese classic literature.

V. Results and Discussion

A. Tool usability

From the overall 10 sessions, only 5 participants (3 from Aveiro and 2 from Lisbon) have fully completed the recording sessions. The other subjects, even though they passed the preliminary testing, showed major difficulties in carrying out or completing the data collection task due to various problems associated to advanced age (such as poor sight capability), low literacy and reduced scholar level, fatigue, difficulty in keeping focused for a relatively long period of time. It should be noted that the 2 subjects from Lisboa were women with the highest level of qualification (the remaining subjects only had Primary education), and cultural level above the other subjects (with only primary school education level). Two other subjects required the Windows Magnifier tool support to be able to complete this task. It should be noted though that using the Magnifier tool slowed down the process, resulting in extreme tiredness of the participant after completing the task, added to the fact that the subject was 92 years old.

In brief, Doaravoz 1.0 speech data collection tool showed several usability issues in this age group, namely:

a) Font size was too small. The majority of the subjects had problems in reading the proposed prompts due to the used small font size. Windows Magnifier was used in order to enlarge certain parts of the screen. This added extra time cost to the session.

b) Control buttons were not positioned correctly, which again turned to be more time consuming because the user had to keep switching between prompts.

c) There were too many instructions. The data collection tool shows several on-screen instructions which didn't prove to be more effective. Subjects often started to read the instructions instead of the actual text to read.

d) Some prompts' content was inappropriate for this age group because: 1) some prompts were too long and used complex words, 2) some sentences contained sensitive words like "death" or family related terms, leading to an immediate emotional change in the user.

e) The end of the speech detection algorithm was incorrectly tuned, due to the naturally slower speech rate of the elderly. This caused some extra repetitions and increased sessions' time.

f) The session proved to be too long (around 1h). The previous issues caused each trial session to get longer, which caused fatigue and boredom in the subjects.

B. Speech recognition of the elderly speech collected data

In order to make a first diagnosis of the existent SR systems' performance in European Portuguese when dealing with elderly speech, a comparison test between two commercial SR systems was conducted. The experiment was done using the speech corpus that was collected from both senior groups.

The procedure was as follows: a) all valid sessions from the database and corresponding utterances were extracted; b) all unique words from all utterances were extracted and analyzed and a word list was built; c) for each utterance it was created a simple, one rule loop CFG grammar containing: all words contained in the prompt (**N**), **N** random words extracted from the complete word list; d) the SR engine was fed with recorded utterance and specific grammar for that utterance; e) the recognition result was extracted and compared with the original prompt (note that the utterances haven't been transcribed; we assumed that the original prompts were correctly read by the speaker); f) the substitutions, deletions and insertions were calculated (as depicted in Table I); g) a recognition result that yielded a null value was considered a full prompt deletion; h) based on the total number of substitutions, deletions and insertions, the global WER was calculated using the formula: WER =

(S+D+I) / TotalWords. From the results shown in Table I, we can see that both SR systems still present a very high WER (36% and 28%), considering that the reference WER for English in Real Time Speech Recognition is less than 16% [42].

Table I: SR Engines comparison test results.

Engine	Substitutions	Deletions	Insertions	Total words	WER
SR Engine 1	269	492	66	2310	36%
SR Engine 2	103	498	50	2310	28%

Further tests need to be performed with the corpus correctly transcribed and with more sessions and utterances. Nevertheless, this test proved the need to collect elderly speech data in order to train the acoustic models with it and improve the SR performance.

VI. Improved Data Collection tool for the Elderly

Based on the results presented in previous section, an updated version of the data collection tool designed for seniors was produced and is already available online (http://www.doaravoz.com) (Fig. 3), by the time of submission of this article.

The following actions were taken in the latest version:

1) Font size was enlarged. The majority of the subjects showed several difficulties in reading the initial prompts due to the small font size.

2) The position of some control and navigation buttons was changed according with users' feedback to a more intuitive and logic position.

3) Instructions were simplified and made more clear using images. The tool contained several text instructions that only increased confusion in the subjects. Often, the subjects tried to read the instruction text itself instead of the actual text to read. The graphical user interface was simplified and more clear instructions were provided, using images.

4) Utterances were simplified. The prompts were too complex and difficult from a syntactic and semantic point of view. **Content was filtered.** The prompts contained sensitive topics such as references to "death" or "dead relatives". These words or references were filtered.

5) Speech detection algorithm was adapted. The speech detection algorithm was often prematurely triggered due to the paused and hesitant nature of elderly speech. The algorithm was adapted to deal with longer pauses that are normal in the seniors' interaction with the system.

6) The session was designed to be conducted in 30-40 minutes. The number of prompts per session was reduced and the instructions simplified.

VII. Conclusions

The following conclusions can be derived from this pilot experiment and can be extended to web and software-based interfaces in general:

a) Education level plays a significant role in the way the user interacts with computers. In the Lisbon sample group, all subjects showed fewer difficulties in completing the recordings. Subjects from the Aveiro group took more time to understand the purpose of the test, as they weren't familiar with the used technology. Therefore, a data collection task of this kind should be performed by an elderly population group with a higher literacy level.

b) Elderly users need more time to deal with tools and technology. This must be considered when programming the time for the users interaction with the systems.

c) The data collection tool required several changes in order to deal with senior users. Main changes included design, font and content changes and other technical aspects.

d) The existent commercial SR systems have a very high WER when dealing with elderly speech, because they are not adapted to seniors' voices. Large amounts of senior speech data are required to improve SR accuracy and to allow Assistive Technology to make use of speech interfaces.

The speech data collection campaign has already started and is on progress. So far, 13 seniors from 60-80 years old have donated their voice. A preliminary usability study with the first 10 was conducted with very positive feedback. Nevertheless, a continuous improvement to the platform is envisioned.

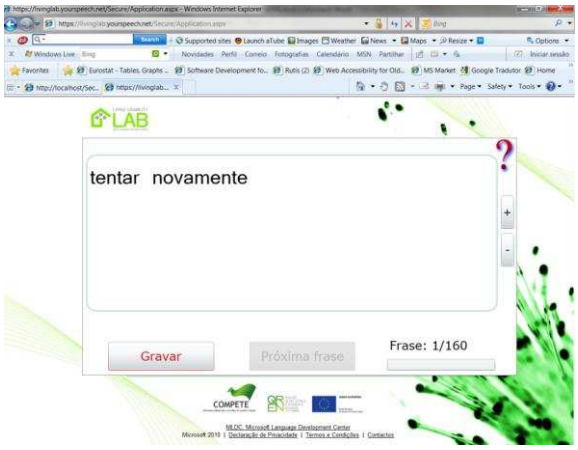

Fig. 3. Screenshot of the Doaravoz 2.0.

REFERENCES

[1] Teixeira, A., Braga, D., Coelho, L., Fonseca, J., Alvarelhão, J., Martín, I., Queirós, A., Rocha, N., Calado, A., Dias, M. "Speech as the Basic Interface for Assistive Technology", DSAI 2009 - Proceedings of the 2th International Conference on Software Development for Enhancing Accessibility and Fighting Info-Exclusion, Porto Salvo, Portugal, 2009.

[2] Neto, N., Patrick, C., Adami, A.G., Klautau, A. G., "Spoltech and ogi-22 baseline systems for speech recognition in Brazilian portuguese", in Teixeira, A., Lima, V., Oliveira, L., Quaresma, P. (eds) Propor 2008, LNCS (LNAI), vol. 5190, pp. 256-259, Springer, Heidelberg, 2008.

[3] Ryan, E. & Cole, R. Evaluative perceptions of interpersonal communication with elders. In: Giles, H., Coupland, N. & Wiemann, J. M. (eds.): Communication, health and the elderly. London: Manchester University Press, pp. 172-191, 1990.

[4] Linville, S.E.: Vocal Aging. Singular, San Diego, 2001.

[5] Jessen, M.: Speaker Classification in Forensic Phonetics and Acoustics. In: Müller, C. (ed.) Speaker Classification I. LNCS(LNAI), vol. 4343, Springer, Heidelberg, 2007.

[6] Wilpon, J. G., and Jacobsen, C. N., "A Study of Speech Recognition for Children and the Elderly", IEEE International Conference on Acoustics, Speech, and Signal Processing. Atlanta, p. 349, 1996.

[7] Fiehler, R. „Kommunikation im Alter und ihre sprachwissenschaftliche Analyse. Gibt es einen Kommunikationsstil des Alters?", Sandig, B. & Selting, M. (Hrsg.): Sprech- und Gesprächsstile. Berlin/New York: de Gruyter, pp. 345-370, 1997.

[8] Helfrich, H. "Age markers in speech", Scherer, K. & Giles, H.: Social markers in speech. Cambridge: University Press, 1979.

[9] Gerritson, M. (1985): Alters- und geschlechtsspezifische Sprachverwendung. In: Besch, W. & Mattheier, K. J. (eds.): Ortssprachenforschung. Beiträge zu einem Bonner Kolloquium. Berlin: Schmidt; 79-108.

[10] Stover, S. E. & Haynes, W. O. "Topic manipulation and cohesive adequacy in conversations of normal adults between the ages of 30 and 90" Clinical Linguistics & Phonetics, 3: 137-149, 1989.

[11] Kemper, S. "Life-span changes in syntactic complexity", Journal of Gerontology, 42: 323-328, 1987.

[12] Schultz, T. "Speaker Characteristics", C. Müller (Ed.) Speaker Classification, Lecture Notes in Computer Science / Artificial Intelligence, Springer, Heidelberg - Berlin - New York, Volume 4343, 2007. To appear.

[13] Xue, S.A., Hao, G.J., "Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study", Journal of Speech, Language, and Hearing Research 46, 689–701, 2003.

[14] Baeckman, L., Small, B. J., & Wahlin, A, "Aging and memory: Cognitive and biological perspectives", In J. E. Birren & K. W. Schaie (Eds.), Handbook of the psychology of aging (pp. 349–377). San Diego, CA, USA: Academic Press, 2001.

[15] Fozard, J.L., & Gordon-Salant, S. "Changes in vision and hearing with aging", J. E. Birren & K. W. Schaie (Eds.), Handbook of the psychology of aging (pp. 241–266). San Diego, CA, USA: Academic Press, 2001.

[16] Wilpon, J. G., and Jacobsen, C. N., "A Study of Speech Recognition for Children and the Elderly", IEEE International Conference on Acoustics, Speech, and Signal Processing. Atlanta, p. 349,1996.

[17] Anderson, S., Liberman, N., Bernstein, E., Foster, S., Cate, E., Levin, B., and Hudson, R., "Recognition of elderly speech and voice-driven document retrieval", Proceedings of the Acoustics, Speech, and Signal Processing, 1999, on 1999 IEEE international Conference - Volume 01 (March 15 - 19, 1999).

[18] Akira, B.; Shin'ichi, Y., Miichi, Y., Lee A., Kiyohiro, S. "Elderly Acoustic Models for Large Vocabulary Continuous Speech Recognition" IEICE Transactions on Information and Systems, Pt.2 (Japanese Edition), 0915-1923, Vol.J85-D-2; No.3; Page 390-397, 2002.

[19] Vipperla, R., Wolters, M., Georgila, K., and Renals, S. "Speech Input from Older Users in Smart Environments: Challenges and Perspectives", Universal Access in Human-

Computer Interaction. Intelligent and Ubiquitous Interaction Environments, Lecture Notes In Computer Science, vol. 5615. Springer-Verlag, Berlin, Heidelberg, 2009.
[20] Fox, S. "Truth about Wired Seniors - Are Wired Seniors Sitting Ducks?", Pew Internet & American Life, 2006.
[21] I2Net, http://www.i2net.com.au/Welcome_.html, last visited on 26-03-2010.
[22] Hönig, Florian; Hacker, Christian; Warnke, Volker; Nöth, Elmar; Hornegger, Joachim; Kornhuber, Johannes Developing Enabling Technologies for Ambient Assisted Living: Natural Language Interfaces, Automatic Focus Detection and User State Recognition In: BMBF (Bundesministerium für Bildung und Forschung); Tagungsband zum 1. deutschen AAL-Kongress (1. Deutscher AAL (Ambient Assisted Living)-Kongress Berlin 30.01.2008-01.02.2008) Berlin/Offenbach : VDE Verlag GMBH 2008, pp. 371-375 - ISBN 978-3-8007-3076-6
[23] QualiWorld, http://qualilife.com/products/index.cfm?id=191&prodType=0&prodTarget=6, (08-04-2010).
[24] NIH Senior Health website, http://nihseniorhealth.gov/, (08-06-2010)
[25] Verbose Text to Speech, http://www.nch.com.au/verbose/index.html, (26-03-2010).
[26] "Census 2001: Análise da população com deficiência", Instituto Nacional de Estatística, in : http://www.inr.pt/download.php?filename=Censos+2001&file=%2Fuploads%2Fdocs%2Finfestatistica%2FCensos2001Destaque.pdf (28/01/2010)
[27] "Projecções de População Residente, Portugal e NUTS II 2000-2050", Instituto Nacional de Estatística (2004), in http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=96942&att_display=n&att_download=y (06/07/2010)
[28] Source: "Estatísticas Demográficas 2008", Instituto Nacional de Estatística (2009), in http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=78180328&att_display=n&att_download=y (06/07/2010)
[29] Source: "Anuário Estatístico de Portugal 2008", Instituto Nacional de Estatítica (2009), in http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=83354386&att_display=n&att_download=y (5/07/2010)
[30] "Revista de Estudos Demográficos - Nº 40", Instituto Nacional de Estatística (2007), in http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=5993008&att_display=n&att_download=y (04/07/2010)
[31] Re-captcha website, http://recaptcha.net/, (20/042010).
[32] Google new account creation, http://mail.google.com/mail/signup (20/04/2010).
[33] Tag a Tune, Online: http://www.gwap.com/gwap/gamesPreview/tagatune/ (20/04/2010)..
[34] Galaxy zoo, http://www.galaxyzoo.org/ (20/04/2010).
[35] Idea Bounty, http://www.ideabounty.com/ (20/04/2010)..
[36] Google Image labeler, http://images.google.com/imagelabeler/ (20/04/2010).
[37] PopVideo,http://www.gwap.com/gwap/gamesPreview/popvideo/ (20/04/2010).
[38] Wikipedia.org,: http://en.wikipedia.org/wiki/Wikipedia (20/04/2010).
[39] João Freitas, António Calado, Pedro Silva, Bruno Reis, Daniela Braga, Miguel Dias, "YourSpeech: Desktop speech data collection based on crowd sourcing in the Internet", PROPOR 2010, 28-30 April, 2010.
[40] Brabham, D. C., "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases", in: Convergence The International Journal of Research into New Media Technologies, Vol. 14(1), pp. 75-90, 2008.
[41] SpeechDat website: http://www.speechdat.org/SpeechDat.html (30/07/2010)
[42] Matsoukas, S., Prasad, R., Laxminarayan, S., Xiang, B., Nguyen, L., Schwartz, R., "The 2004 BBN 1xRT Recognition Systems for English Broadcast News and Conversational Telephone Speech", Interspeech 2005, Lisbon, Portugal, 2005.