The EASR Corpora of European Portuguese, French, Hungarian and Polish Elderly Speech

Annika Hämäläinen[1,2], Jairo Avelar[1], Silvia Rodrigues[1], Miguel Sales Dias[1,2], Artur Kolesiński[3], Tibor Fegyó[4], Géza Németh[4], Petra Csobánka[5], Karine Lan Hing Ting[6] and David Hewson[6]

[1]Microsoft Language Development Center
[2]ISCTE – University Institute of Lisbon
[3]Knowledge Society Association
[4]Department of Telecommunications & Media Informatics, Budapest University of Technology & Economics
[5]Institute for Infocommunication Technologies, Bay Zoltán Nonprofit Ltd. for Applied Research
[6]Troyes University of Technology
[1]Rua do Fogo de Santelmo, Lote 2.07.02, 1990-110 Lisbon, Portugal
[3]ul. Grażyny 13/15 lok. 221, 02-548 Warsaw, Poland
[4]Magyar tudósok krt. 2., 1117 Budapest, Hungary
[5]Fehérvári út 130, 1116 Budapest, Hungary
[6]12 rue Marie Curie, CS 42060, 10004 Troyes, France
E-mail: {t-anhama, t-jairol, miguel.dias}@microsoft.com, artur.kolesinski@ssw.org.pl, {fegyo, nemeth}@tmit.bme.hu, petra.csobanka@bayzoltan.hu, {karine.lan, david.hewson}@utt.fr

Abstract

Currently available speech recognisers do not usually work well with elderly speech. This is because several characteristics of speech (e.g. fundamental frequency, jitter, shimmer and harmonic noise ratio) change with age and because the acoustic models used by speech recognisers are typically trained with speech collected from younger adults only. To develop speech-driven applications capable of successfully recognising elderly speech, this type of speech data is needed for training acoustic models from scratch or for adapting acoustic models trained with younger adults' speech. However, the availability of suitable elderly speech corpora is still very limited. This paper describes an ongoing project to design, collect, transcribe and annotate large elderly speech corpora for four European languages: Portuguese, French, Hungarian and Polish. The Portuguese, French and Polish corpora contain read speech only, whereas the Hungarian corpus also contains spontaneous command and control type of speech. Depending on the language in question, the corpora contain 76 to 205 hours of speech collected from 328 to 986 speakers aged 60 and over. The final corpora will come with manually verified orthographic transcriptions, as well as annotations for filled pauses, noises and damaged words.

Keywords: automatic speech recognition, corpus, elderly speech

1. Introduction

The proportion of the elderly (people aged 65 and over) is growing fast in European countries (Eurostat, 2011). This demographic change calls for rapid action to establish "active aging" policies and programmes that help older adults remain healthy and active and live in their own homes as long as possible (World Health Organization, 2002). Information and communication technology (ICT) has considerable potential when it comes to facilitating the lives of the elderly. Any applications developed for older adults must, however, take their special needs and preferences into account. As the elderly often have difficulties using small devices and complex interfaces, the naturalness and ease of speech interfaces would be particularly interesting in their case (Teixeira et al., 2012; Teixeira et al., 2009). However, currently available automatic speech recognisers are typically trained using younger adults' speech and do not work well with elderly speech. This is because several characteristics of speech (e.g. fundamental frequency, jitter, shimmer and harmonic noise ratio) change with age (Xue & Hao, 2003). For instance, Wilpon & Jacobsen (1996) and Pellegrini et al. (2012) showed that, as compared with speech recognition results achieved on younger adults' speech, the increase in word error rate is considerable when recognising elderly speech (in this case, read speech from Danish subjects aged over 70 and Portuguese subjects aged 60 and over, respectively) using acoustic models (AMs) trained with speech collected from younger adults. On the other hand, in the experiments of Anderson et al. (1999), recognising American English elderly speech (both spontaneous and read) using elderly-specific AMs led to a significant decrease (about 12.5% absolute) in WER as compared with AMs trained using speech collected from younger adults. These results show that there is a real need for adapting AMs to elderly speech.

To develop speech-driven applications capable of successfully recognising elderly speech, a sufficient amount of this type of data is needed for training the AMs. The largest resource of elderly speech is the MALACH (Multilingual Access to Large Spoken Archives) database (see Section 7), which contains 160.000 hours of interviews in 32 languages (e.g. English, Hungarian) from survivors, liberators, rescuers and witnesses of the Holocaust. However, the elderly speech in the MALACH database is atypical in the sense that much of it is non-native and/or emotional speech. Elderly speech corpora that are better suited for automatic speech recognition (ASR) purposes include, for instance, the 25-hour Jasmin-CGN corpus of Dutch and Flemish elderly speech (Cucchiarini et al., 2006) and the 133-hour S-JNAS corpus of Japanese elderly speech (Baba et al., 2002). However, in the case of most languages, the availability of suitable elderly speech corpora is still very limited.

This paper describes the work done to develop elderly speech corpora for four European languages that are currently under-resourced when it comes to the availability of elderly speech corpora suitable for ASR purposes: Portuguese, French, Hungarian and Polish. The corpora are called the EASR (Elderly Automatic Speech Recognition) Corpora of European Portuguese (Hämäläinen et al., 2012), French, Hungarian and Polish Elderly Speech. They are being developed in four publicly funded projects developing speech-driven applications for the elderly. The development of the Portuguese corpus started in the QREN Living Usability Lab (LUL) project (see Section 7), and has continued in the QREN Smartphones for Seniors (S4S; see Section 7) and QREN FalaGlobal projects, while all the other corpora are being developed in the Personal Assistant to Enhance the Social Life of Seniors (PaeLife) project (see Section 7). All of these projects aim at developing applications to help the elderly stay healthy, active, independent and socially connected. The Portuguese data collection was organised by Microsoft Language Development Center (MLDC), whereas the French, Hungarian and Polish data collections are a collaboration between MLDC and the PaeLife partners in France (Troyes University of Technology), Hungary (Budapest University of Technology & Economics, and Bay Zoltán Nonprofit Ltd. for Applied Research), and Poland (Knowledge Society Association).

The EASR Corpora contain clean wideband speech read out by native speakers aged 60 and over – that is, 16-bit, 16/22-kHz elderly read speech collected in a quiet environment using a laptop and a USB microphone headset. In addition to read speech, the Hungarian corpus contains spontaneous command and control type of speech.

2. Corpus Design

The goal of the data collection was to develop sufficiently large corpora of elderly speech for training/adapting and testing AMs for the speech-driven parts of the applications developed in the LUL, S4S, FalaGlobal and PaeLife projects, as well as for acoustic-phonetic research on elderly speech. The following subsections describe the selection of speakers and speaker prompts for the corpora.

2.1 Speaker Selection

There do not seem to be clear-cut age-related acoustic-phonetic differences that would set elderly speech apart from younger adults' speech (Ryan & Cole, 1990); the aging of speech is not only affected by chronological age but people's lifestyles (e.g. smoking, alcohol consumption, psychological stress, abuse and overuse of vocal cords) (Linville, 2001), as well. While it might be impossible to determine the exact age from which an individual's speech could be considered elderly, studies usually regard 60-70 years of age as the minimum age range for elderly speech (Wilpon & Jacobsen, 1996). That is why we collected speech from speakers aged 60 and over for the EASR corpora.

To be able to read the speaker prompts presented to them using a web interface specifically adapted to the elderly (see Section 3), the speakers are required to have good enough eyesight and reading and technical comprehension skills. We have been able to source suitable speakers from 3rd age universities, as well as care institutions, social clubs and/or associations for seniors located in different parts of Portugal, France, Hungary and Poland, for a variety of regional accents.

Other than the criteria mentioned above, we do not have stringent criteria for selecting the elderly speakers. We do not, for instance, aim at a specific ratio of female and male speakers or a specific ratio of speakers representing different regional accents.

2.2 Prompt Selection

When designing the material to be read out by the elderly speakers, our goal was to design a set of prompts that would be suitable for training automatic speech recognisers for a wide variety of speech-driven applications ranging from command and control to dictation applications. In the case of Portuguese and Hungarian, the prompts include phonetically rich sentences, as well as command and control type of prompts (e.g. *show me the next video*), whereas the French and Polish corpora only contain phonetically rich sentences. The differences in the design (and size) of the four corpora are due to differences in the resources and expertise available for the work in the four countries.

2.2.1. Portuguese

The design of the Portuguese prompts was inspired by existing Portuguese corpora such as SpeechDat II (see Section 7). The prompts included phonetically rich sentences and common types of prompts used in speech-driven applications (e.g. isolated digits, numbers, dates, times). Additionally, prompts specific to the speech-driven applications developed for Ambient Assisted Living scenarios in the LUL project (e.g. Living Home Center (Pires et al., 2012)) were included. Table 1 illustrates the number of unique prompts per prompt type, as well as the number of different types of prompts uttered by each speaker. Application phrases contained full commands considered relevant for the LUL applications, which include, for instance, speech-enabled email (e.g. *abrir mensagem* ('open message')) and calendar (e.g. *activar alarme* ('activate alarm')). Application words were shorter versions of application phrases (e.g. *responder* ('respond'), *favoritos* ('favourites')), and web words consisted of important web sites (e.g. *Hotmail*), applications (e.g. *Internet Explorer*) and key strikes (e.g. *enter*). PIM refers to commands that might be used in Personal Information Management applications with commands for meeting request management, calendar management etc. The other prompt types are self-explanatory and only described in more detail in Hämäläinen et al. (2012).

Each speaker was asked to utter a set of 160 prompts selected across the 14 different prompt types (see Table 1). Uttering 160 prompts resulted in about 7 minutes of recordings per speaker.

Prompt Type	#Unique Prompts	Prompts Per Speaker
Phonetically Rich Sentences	3497	60
Natural Numbers	750	5
Isolated Digits	10	2
Phone Numbers	150	10
PIN Codes	150	5
Dates	750	10
Times	1125	10
Time Ranges	370	5
Person Names	750	10
City Names	500	15
Application Words	229	10
Application Phrases	416	10
Web Words	29	5
PIM	210	3

Table 1: The different prompt types appearing in the Portuguese corpus.

	#Word Types	#Word Tokens	Audio/Hours
Portuguese	10,400	810,000	188
French	10,500	405,000	76
Hungarian	37,000	500,000	113
Polish	18,000	815,000	205

Table 2: The main statistics of the corpora at the time of writing this paper.

2.2.2. French

The French prompts only included phonetically rich sentences. Each French speaker was asked to utter a set of 160 prompts that were selected randomly from a pool of 10.639 different phonetically rich sentences. Uttering 160 prompts resulted in about 14 minutes of recordings per speaker.

2.2.3. Hungarian

The Hungarian prompts include phonetically rich sentences, as well as prompts designed to elicit spontaneous command and control type of speech. Each speaker is asked to utter a set of 60 phonetically rich sentences selected randomly from a pool of 18.922 different phonetically rich sentences. In addition, each speaker produces 20 utterances of spontaneous command and control type of speech (e.g. *I would like to listen to some sonatas of Bach; Please read the news to me; Please call the doctor*), which is elicited using a set of 76 unique prompts requesting the speakers to ask a virtual assistant to do something for them. The prompt design used for Hungarian results in about 10 minutes of recordings per speaker.

2.2.4. Polish

The Polish prompts only included phonetically rich sentences. Each Polish speaker was asked to utter a set of 160 prompts that were selected randomly from a pool of 10.416 different phonetically rich sentences. Uttering 160 prompts resulted in about 16 minutes of recordings per speaker.

3. Data Collection

We use the web-based *Microsoft Your Speech* data collection platform (Freitas et al, 2010) for collecting the speech and some biographical information about the speakers (age group, gender and the region they had grown up in). The platform's web interface was localised for each of the four languages and specifically adapted for elderly users, for example, by using a large font size and by calibrating the speech detection algorithm such that it allows longer pauses, which are characteristic of elderly speech (Helfrich, 1979). The recording sessions are supervised by people trained for the task and take place in quiet rooms. The web interface is operated either by the recording supervisor or by the speaker – whichever the speaker is more comfortable with. Each utterance is recorded separately using a noise-cancelling Life Chat LX 3000 USB headset and sampled at 16 bits and 22 kHz. In the Portuguese corpus, part of the data was sampled at 16 bits and 16 kHz.

The recording of an utterance starts when the *Record* button is clicked and ends when the *Stop* button is clicked or when no more speech is detected by the system. The recorded utterance is then uploaded to the web backend of the system and checked for the presence of speech and clipping. If speech is indeed detected and the utterance does not contain any samples of clipping, the user may proceed to the next prompt by clicking the *Next prompt* button, or repeat the utterance by clicking the *Rerecord* button. If the automatic quality control is not passed, the speaker is requested to re-record the utterance.

4. The EASR Corpora

4.1 Overview of the Speech Data

The data collection effort has already been completed for Portuguese, French and Polish but is still ongoing in the case of Hungarian. Table 2 shows the approximate amount of speech collected for each language in terms of the number of word types and tokens, as well as hours of recorded audio. The Portuguese, French and Polish corpora contain about 188, 76 and 205 hours of read speech from 986, 328 and 781 elderly speakers, respectively. At the time of writing this paper, we had already collected about 113 hours of speech from 678 Hungarian speakers; we intend to collect as much Hungarian elderly speech as is possible within the time frame and budget available in the PaeLife project. There are more female than male speakers in all four corpora: 72% of the speakers in the Portuguese corpus, 73% of the speakers in the French corpus, 81% of the speakers in the Polish corpus and, at the moment, 76% of the speakers in the Hungarian corpus are female. In our experience, elderly females are more willing to contribute to speech data collection campaigns than elderly males. Tables 3-6 present information about the age distribution and the regional origin of the speakers in each corpus. As one might expect, the number of subjects per age group decreases with age. Demographically speaking, the collected speech data might not be fully representative of the elderly population of the countries in question but we have still managed to collect speech from a large number of elderly speakers in each country (e.g. S-JNAS only collected speech from 301 elderly speakers (Baba et al., 2002)).

4.2 Transcriptions and Annotations

As already mentioned, the Portuguese, French and Polish corpora only contain read speech, whereas the Hungarian corpora also contains spontaneous command and control type of speech. The read speech in all corpora is transcribed orthographically by using the speaker prompts as the starting point for the transcription work. Before manual verification of these initial transcriptions, we use text normalisation rules to convert all number expressions (e.g. dates, times) and abbreviations into their most likely expanded forms, and to remove all punctuation (the only exceptions are word-internal punctuation and spelled letters, which are disambiguated from one-letter words using a trailing period, e.g. "*A.*"). Native speakers then listen to the recordings and, when necessary, edit the initial transcriptions such that they correspond to what the speakers are saying in the recordings. The spontaneous speech in the Hungarian corpus is transcribed (orthographically) from scratch.

While verifying or producing the orthographic transcriptions, the transcribers also annotate filled pauses, noises, damaged words (e.g. mispronunciations, false starts), and speech from non-primary speakers (e.g. the recording supervisor) using the tags in Table 7. The annotation scheme was designed to be compatible with the requirements of the in-house tools that we use for training our AMs.

We are using two different methods for the transcription and annotation work: 1) hiring and training native speakers to do the work using an in-house transcription tool and 2) using a third-party vendor to hire a crowd of native speakers to do the work online using the three-step crowdsourcing method described by Hämäläinen et al. (2013).

The transcription and annotation of all four corpora is either underway or about to start. About 75% of the Portuguese data and about 10% of the Hungarian data have already been transcribed and annotated. The transcription and annotation work is about to start for French and Polish.

Tag	Meaning
<FILL/>	Filled pauses (e.g. "umm", "er", "ah")
<NON/>	Non-human noises (e.g. mouse or keyboard noises, radio, TV, music)
<SPN/>	Human noises (e.g. coughs, audible breath, prompt echoes)
<UNKNOWN/>	Unintelligible and truncated words, unclear speech from non-primary speakers
<NPS/>	Speech from non-primary speakers

Table 7: The tags available for annotating the elderly speech corpora.

Age/Area	Centre	South	North	Madeira	Azores	Abroad	Total
60-65	240	21	128	1	0	10	400
66-70	106	14	47	2	0	12	181
71-75	86	15	50	2	1	6	160
76-80	55	6	25	0	0	5	91
81-85	32	3	18	0	0	0	53
86-90	18	1	12	0	0	1	32
91-95	0	0	0	0	0	0	0
96-100	0	0	1	0	0	0	1
Unknown	45	1	18	0	0	4	68
Total	582	61	299	5	1	38	986

Table 3: The number of Portuguese speakers per age group and regional origin. "Unknown" contains speakers aged 60 and over whose exact age group was not recorded.

Age/Area	Alsace	Aquitaine	Auvergne	Bretagne	Bourgogne	Centre	Champagne-Ardenne	Franche-Comté	Île-de-France	Languedoc-Roussillon	Limousin
60-64	3	1	0	0	11	0	85	0	7	0	1
65-69	3	0	1	0	3	0	26	0	3	0	0
70-74	0	0	0	0	0	1	24	1	1	0	0
75-79	1	0	0	0	1	1	20	0	2	0	0
80-100	1	0	0	0	2	0	76	0	2	2	0
Unknown	3	0	0	0	0	0	4	0	0	0	0
Total (All Ages)	11	1	1	0	17	2	235	1	15	2	1
Age/Area	**Lorraine**	**Basse-Normandie**	**Midi-Pyrénées**	**Nord-Pas-de-Calais**	**Pays de la Loire**	**Picardie**	**Poitou-Charentes**	**Provence-Alpes-Côte d'Azur**	**Rhône-Alpes**	**Haute-Normandie**	**Total (All Areas)**
60-64	1	0	1	3	1	7	0	0	1	0	122
65-69	2	0	0	1	1	1	0	0	1	0	42
70-74	2	1	0	1	0	2	0	0	0	0	33
75-79	0	0	0	1	0	5	0	2	1	0	34
80-100	1	0	0	1	0	4	0	0	1	0	90
Unknown	0	0	0	0	0	0	0	0	0	0	0
Total (All Ages)	6	1	1	7	2	19	0	2	4	0	328

Table 4: The number of French speakers per age group and regional origin.

Age/Area	Nyugat-Dunántúl	Dél-Dunántúl	Közép-Dunántúl	Dél-Alföld	Észak-Alföld	Észak-Magyarország	Budapest	Felvidék (Szlovákia)
60-64	17	3	41	8	5	39	46	1
65-69	69	4	23	3	4	26	31	1
70-74	24	10	28	9	6	5	20	0
75-79	24	15	16	11	4	2	11	2
80-100	37	24	26	12	13	9	39	3
Total (All Ages)	171	56	134	43	32	81	147	7
Age/Area	**Kárpátalja (Ukrajna)**	**Partium (Románia)**	**Erdély (Románia)**	**Vlachia-Moldva (Románia)**	**Vajdaság (Szerbia)**	**Burgenland (Ausztria)**	**Egyéb**	**Total (All Areas)**
60-64	0	1	1	0	0	0	0	162
65-69	0	0	1	0	0	0	0	162
70-74	0	0	0	0	0	0	0	102
75-79	0	0	1	0	0	0	1	87
80-100	0	0	1	0	0	0	1	165
Total (All Ages)	0	1	4	0	0	0	2	678

Table 5: The number of Hungarian speakers per age group and regional origin, at the time of writing this paper.

Age/Area	Dolnoslaskie	Kujawsko-Pomorskie	Lódzkie	Lubelskie	Malopolskie	Mazowieckie	Opolskie	Podkarpackie
60-64	5	66	16	2	11	203	0	1
65-69	10	58	28	0	10	118	0	1
70-74	3	43	6	0	4	60	0	0
75-79	3	21	4	1	1	38	0	0
80-100	0	21	7	0	1	21	0	0
Total (All Ages)	21	209	61	3	27	440	0	2

Age/Area	Podlaskie	Pomorskie	Slaskie	Swietokrzyskie	Warminsko-Mazurskie	Wielkopolskie	Zachodniopomorskie	Total (All Areas)
60-64	2	0	2	2	0	0	1	311
65-69	0	2	0	1	0	0	0	228
70-74	1	0	1	0	0	1	0	119
75-79	0	0	1	1	0	1	0	71
80-100	1	0	1	0	0	0	0	52
Total (All Ages)	4	2	5	4	0	2	1	781

Table 6: The number of Polish speakers per age group and regional origin.

4.3 Quality Control

As mentioned in Section 3, the data collection platform performs some instant data quality checking. By using the platform, we can be sure that our recordings indeed contain speech, and that they do not contain any samples of clipping. In addition, we use forced alignment to identify potentially problematic utterances; utterances that cannot be aligned are likely to have problems in their orthographic transcriptions and/or their audio quality. If possible, the problems are fixed; if not, the utterances are removed from the corpora.

During the transcription and annotation work, the transcribers are able to discard sessions that contain recordings with consistently poor audio quality, or speech from non-native speakers or speakers with consistent problems reading the prompts. For this reason, the figures presented in Tables 2-6 are subject to change.

To identify errors in the orthographic transcriptions of words, as well as “new” words that we do not have in our pronunciation lexica, we break the orthographic transcriptions into individual words, and check those words against the pronunciation lexicon of the language in question. Using this simple method, we are able to identify misplaced tags (e.g. *vi<SPN/>rgula* (‘*co<SPN/>mma*’)), words that need to be added into the pronunciation lexicon for ASR purposes, as well as words with typos. Some of the typos can be corrected automatically (e.g. *irgula* is bound to refer to *virgula* in the Portuguese corpus), while ambiguous typos are checked against the recordings and corrected manually (e.g. *ete* could be a misspelling of several different words in Portuguese).

5. Summary

In this paper, we described the design, collection, transcription and annotation of four large corpora of elderly read speech collected from native speakers of European Portuguese, French, Hungarian and Polish who are aged 60 and over. At the time of writing this paper, these EASR Corpora of European Portuguese, French, Hungarian and Polish Elderly Speech contained 188, 76, 113 and 205 hours of speech (a total of 582 hours of speech) recorded from 986, 328, 678 and 781 elderly speakers (a total of 2773 elderly speakers), respectively. The Hungarian data collection is still ongoing. The transcription and annotation of the corpora is underway; the final corpora will come with manually verified orthographic transcriptions, as well as annotations indicating filled pauses, noises, damaged words (e.g. mispronunciations) and speech from non-primary speakers. The corpora will play an important role in ensuring that speech-driven, easy-to-use technology can be made available to the growing elderly populations of Portugal, France, Hungary and Poland. The corpora will also be useful for acoustic-phonetic research on the characteristics of elderly speech (see Pellegrini et al. (2013) for an analysis of the elderly speech in the EASR Corpus of European Portuguese Elderly Speech). The Portuguese, French, and Polish corpora will be available at request for R&D activities and will soon be published in well-known catalogues. Please contact Miguel Sales Dias (miguel.dias@microsoft.com) for further information about the Portuguese corpus, David Hewson (david.hewson@utt.fr) for further information about the French corpus and Artur Kolesiński (artur.kolesinski@ssw.org.pl) for further information about the Polish corpus. The public availability of the Hungarian corpus is currently under discussion.

6. Acknowledgements

This work was partially funded by: (1) the European Community, Fundação de Ciência e Tecnologia (FCT) in Portugal, the National Research Agency (ANR) in France, the Ministry for National Economy (NGM) in Hungary, and the National Centre for Research and Development in Poland, through the AAL program (PaeLife Project N°AAL-2009-2-068); (2) European Structural Funds for Portugal (FEDER) through COMPETE (Operational Program for Competitiveness Factors), as part of the National Strategic Reference Framework (QREN), the national program of incentives for Portuguese businesses and industry, through QREN 21541 Smartphones for Seniors and QREN 7900 Living Usability Lab; (3) FEDER through POR Lisboa (Regional Operational Programme of Lisbon), as part of QREN, through QREN. 5329 FalaGlobal.

7. References

Anderson, S., Liberman, N., Bernstein, E., Foster, S., Cate, E., Levin, B., Hudson, R. (1999). Recognition of elderly speech and voice-driven document retrieval. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Phoenix, AZ, USA.

Baba, A., Yoshizawa, S., Yamada, M., Lee, A., Shikano, K. (2002). Elderly acoustic models for large vocabulary continuous speech recognition. *Transactions of the Institute of Electronics, Information and Communication Engineers*, D-II J85D-II, pp. 390--397.

Cucchiarini, C., Van hamme, H., van Herwijnen, O., Smits, F. (2006). JASMIN-CGN: Extension of the spoken Dutch corpus with speech of elderly people, children and non-natives in the human-machine interaction modality. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, Genoa, Italy.

Eurostat (2011). *Population projections*, http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_projections, last accessed 18 March 2014.

Freitas, J., Calado, A., Braga, D., Silva, P., Dias, M. (2010). Crowdsourcing platform for large-scale speech data collection. In *Proceedings of FALA*. Vigo, Spain.

Helfrich, H. (1979). Age markers in speech. In K. Scherer & H. Giles (Eds.), *Social markers in speech*. Cambridge: Cambridge University Press.

Hämäläinen, A., Pinto, F., Sales Dias, M., Júdice, A., Freitas, J., Pires, C., Teixeira, V., Calado, A., Braga, D. (2012). The first European Portuguese elderly speech corpus. In *Proceedings of IberSPEECH*. Madrid, Spain.

Hämäläinen, A., Pinto Moreira, F., Avelar, J., Braga, D., Sales Dias, M. (2013). Transcribing and annotating speech corpora for speech recognition: A three-step crowdsourcing approach with quality control. In *Proceedings of the Second AAAI Conference on Human Computation and Crowdsourcing*. Palm Springs, CA, USA.

Linville, S.E. (2001). *Vocal aging*. San Diego: Singular.

Living Usability Lab (LUL), http://www.livinglab.pt/, last accessed 13 March 2014.

Pellegrini, T., Trancoso, I., Hämäläinen, A., Calado, A., Sales Dias, M., Braga, D. (2012). Impact of age in ASR for the elderly: Preliminary experiments in European Portuguese. In *Proceedings of IberSPEECH*. Madrid, Spain.

Pellegrini, T., Hämäläinen, A., Boula de Mareüil, P., Tjalve, M., Trancoso, I., Candeias, S., Sales Dias, M., Braga, D. (2013). A corpus-based study of elderly and young speakers of European Portuguese: Acoustic correlates and their impact on speech recognition performance. In *Proceedings of Interspeech*. Lyon, France.

Personal Assistant to Enhance the Social Life of Seniors (PaeLife), http://www.paelife.eu/, last accessed 13 March 2014.

Pires, C., Pinto, F., Teixeira, V., Freitas, J., Dias, M. (2012). Living Home Center – A personal assistant with multimodal interaction for elderly and mobility impaired e-inclusion. In *Proceedings of the International Conference on Computational Processing of Portuguese*. Coimbra, Portugal.

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

Smartphones for Seniors (S4S), http://www.smartphones4seniors.org/, last accessed 13 March 2014.

SpeechDat, http://www.speechdat.org/SpeechDat.html, last accessed 13 March 2014.

Teixeira, A., Braga, D., Coelho, L., Fonseca, J., Alvarelhão, J., Martín, I., Queirós, A., Rocha, N., Calado, A., Dias, M. (2009). Speech as the basic interface for assistive technology. In *Proceedings of the Second International Conference on Software Development for Enhancing Accessibility and Fighting Info-Exclusion.* Porto Salvo, Portugal.

Teixeira, V., Pires, C., Pinto., F., Freitas, J., Sales Dias, M., Mendes Rodrigues, E. (2012). Towards elderly social integration using a multimodal human-computer interface. In *Proceedings of the Second International Living Usability Lab Workshop on AAL Latest Solutions, Trends and Applications*, Vilamoura, Portugal.

The Multilingual Access to Large Spoken Archives Database (MALACH), http://malach.umiacs.umd.edu/, last accessed 13 March 2014.

Wilpon, J.G., Jacobsen, C. N. (1996). A study of speech recognition for children and the elderly. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Atlanta, GA, USA.

World Health Organization (2002). *Active aging: A policy network*. Second United Nations World Assembly on Ageing. Madrid, Spain.

Xue, S.A., Hao, G.J. (2003). 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, pp. 689--701.